92

500.43106X00

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s): Hiroki KANAI

Serial No.: 10/654,996

Filed: September 5, 2003

Title: STORAGE DEVICE CONTROL APPARATUS AND CONTROL METHOD FOR THE STORAGE DEVICE CONTROL APPARATUS

LETTER CLAIMING RIGHT OF PRIORITY

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

September 25, 2003

Sir:

Under the provisions of 35 USC 119 and 37 CFR 1.55, the applicant(s) hereby claim(s) the right of priority based on:

Japanese Patent Application No. 2003-111405
Filed: April 16, 2003

A certified copy of said Japanese Patent Application is attached.

Respectfully submitted,

ANTONELLI, TERRY, STOUT & KRAUS, LLP

Carl I. Brundidge
Registration No. 29,621

CIB/rp
Attachment

日本国特許庁 W1160-01EW
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日
Date of Application: ２００３年　４月１６日

出願番号
Application Number: 特願２００３－１１１４０５

[ST. 10/C]: [ＪＰ２００３－１１１４０５]

出願人
Applicant(s): 株式会社日立製作所

２００３年　８月２６日

特許庁長官
Commissioner,
Japan Patent Office 今井康夫

出証番号　出証特２００３－３０６９６３５

【書類名】　特許願
【整理番号】　HI030082
【提出日】　平成15年 4月16日
【あて先】　特許庁長官殿
【国際特許分類】　G06F 3/06
【発明者】
【住所又は居所】　神奈川県小田原市中里３２２番地２号　株式会社日立製作所ＲＡＩＤシステム事業部内
【氏名】　金井　宏樹
【特許出願人】
【識別番号】　000005108
【氏名又は名称】　株式会社日立製作所
【代理人】
【識別番号】　110000176
【氏名又は名称】　一色国際特許業務法人
【代表者】　一色　健輔
【手数料の表示】
【予納台帳番号】　211868
【納付金額】　21,000円
【提出物件の目録】
【物件名】　明細書　1
【物件名】　図面　1
【物件名】　要約書　1
【プルーフの要否】　要

【書類名】　明細書

【発明の名称】　記憶デバイス制御装置、及び記憶デバイス制御装置の制御方法

【特許請求の範囲】

【請求項１】　データ入出力要求を受信するためのホストインタフェース制御部が形成されたチャネル制御ユニットと、

前記データ入出力要求に応じて、データを記憶するための記憶ボリウムに対する前記データの入出力制御を行うためのディスクインタフェース制御部が形成されたディスク制御ユニットと、

前記データを記憶するためのメモリが形成されたキャッシュメモリユニットと、

前記ホストインタフェース制御部と前記ディスクインタフェース制御部と前記メモリとが形成されたストレージ制御ユニットと

を挿抜可能な装着部と、

前記チャネル制御ユニット、前記ディスク制御ユニット、前記キャッシュメモリユニット、及び前記ストレージ制御ユニットを通信可能に接続する内部接続部と

を備えることを特徴とする記憶デバイス制御装置。

【請求項２】　前記ストレージ制御ユニットの前記メモリには、

前記ストレージ制御ユニットが前記入出力制御を行う前記記憶ボリウムを特定するための情報が記憶され、

前記ディスク制御ユニットが前記入出力制御を行う前記記憶ボリウムを特定するための情報と他の前記ストレージ制御ユニットが前記入出力制御を行う前記記憶ボリウムを特定するための情報との少なくともいずれかが記憶されること

を特徴とする請求項１に記載の記憶デバイス制御装置。

【請求項３】　前記ストレージ制御ユニットの前記メモリは、

前記ストレージ制御ユニットが前記入出力制御を行う前記記憶ボリウムに記憶される前記データを記憶するための第１の記憶領域と、

前記ディスク制御ユニットが前記入出力制御を行う前記記憶ボリウムに記憶される前記データを記憶するための第２の記憶領域とを有し、

前記第１及び前記第２の記憶領域を特定するための情報を記憶すること
を特徴とする請求項１に記載の記憶デバイス制御装置。

【請求項４】　前記ストレージ制御ユニットは、前記他のストレージ制御ユニットとの間で授受されるデータを記憶するための通信バッファを備えること
を特徴とする請求項１に記載の記憶デバイス制御装置。

【請求項５】　データ入出力要求を受信するためのホストインタフェース制御部が形成されたチャネル制御ユニットと、

前記データ入出力要求に応じて、データを記憶するための記憶ボリウムに対する前記データの入出力制御を行うためのディスクインタフェース制御部が形成されたディスク制御ユニットと、

前記データを記憶するためのメモリが形成されたキャッシュメモリユニットと
、

前記ホストインタフェース制御部と前記ディスクインタフェース制御部と前記メモリとが形成されたストレージ制御ユニットと
を挿抜可能な装着部と、

前記チャネル制御ユニット、前記ディスク制御ユニット、前記キャッシュメモリユニット、及び前記ストレージ制御ユニットを通信可能に接続する内部接続部と
を備え、

前記装着部には少なくとも、

前記ストレージ制御ユニットの前記メモリに、

前記データ入出力要求の対象となっている前記記憶ボリウムに対する前記入出力制御を行う前記ユニットを特定するための情報が記憶された、複数の前記ストレージ制御ユニット
が装着される
記憶デバイス制御装置の制御方法であって、

ある前記ストレージ制御ユニットが、前記データ入出力要求を受信するステップと、

前記ストレージ制御ユニットが前記情報を参照し、前記入出力要求の対象とな

っている前記記憶ボリウムに対する前記入出力制御を行う前記ユニットを特定するステップと、

前記入出力制御を行う前記ユニットが前記ストレージ制御ユニットである場合には、前記ストレージ制御ユニットが前記入出力制御を行い、前記入出力制御を行う前記ユニットが前記ストレージ制御ユニットでない場合には、他の前記ストレージ制御ユニットが前記入出力制御を行うステップと

を備えることを特徴とする記憶デバイス制御装置の制御方法。

【請求項６】　前記ストレージ制御ユニットが前記受信する前記データ入出力要求が前記データの読み出し要求であって、前記入出力制御を行う前記ユニットが前記ストレージ制御ユニットでない場合には、

前記他のストレージ制御ユニットが前記入出力制御を行う前記ステップは、

前記ストレージ制御ユニットが、前記読み出し要求を前記他のストレージ制御ユニットに送信するステップと、

前記他のストレージ制御ユニットが、前記読み出し要求に応じて前記入出力制御を行うステップと、

前記ストレージ制御ユニットが、前記他のストレージ制御ユニットから前記データを受信するステップと、

前記ストレージ制御ユニットが、前記受信した前記データを情報処理装置に送信するステップと

であることを特徴とする請求項５に記載の記憶デバイス制御装置の制御方法。

【請求項７】　前記ストレージ制御ユニットが前記受信する前記データ入出力要求が前記データの書き込み要求であって、前記入出力制御を行う前記ユニットが前記ストレージ制御ユニットでない場合には、

前記他のストレージ制御ユニットが前記入出力制御を行う前記ステップは、

前記ストレージ制御ユニットが、前記書き込み要求及び書き込みデータを前記他のストレージ制御ユニットに送信するステップと、

前記他のストレージ制御ユニットが、前記書き込み要求に応じて前記書き込みデータの前記入出力制御を行うステップと

であることを特徴とする請求項５に記載の記憶デバイス制御装置の制御方法。

【請求項８】　請求項７に記載の記憶デバイス制御装置の制御方法において、

前記ストレージ制御ユニットが、前記他のストレージ制御ユニットから前記書き込みデータの前記入出力制御を完了した旨のデータを受信するステップと、

前記ストレージ制御ユニットが、情報処理装置に前記入出力制御を完了した旨のデータを送信するステップと

を備えることを特徴とする記憶デバイス制御装置の制御方法。

【請求項９】　前記ストレージ制御ユニットは、前記他のストレージ制御ユニットとの間で授受されるデータを記憶するための通信バッファを備えており、

前記ストレージ制御ユニットが、前記読み出し要求を前記他のストレージ制御ユニットに送信する前記ステップは、

前記ストレージ制御ユニットが、前記他のストレージ制御ユニットが備える前記通信バッファに前記読み出し要求を書き込むステップと、

前記他のストレージ制御ユニットが、前記他のストレージ制御ユニットの前記通信バッファから前記読み出し要求を読み出すステップとであり、

前記ストレージ制御ユニットが、前記他のストレージ制御ユニットから前記データを受信する前記ステップは、

前記他のストレージ制御ユニットにより前記ストレージ制御ユニットが備える前記通信バッファに書き込まれた前記データを、前記ストレージ制御ユニットが読み出すステップ

であることを特徴とする請求項６に記載の記憶デバイス制御装置の制御方法。

【請求項１０】　前記ストレージ制御ユニットは、前記他のストレージ制御ユニットとの間で授受されるデータを記憶するための通信バッファを備えており、

前記ストレージ制御ユニットが、前記書き込み要求及び書き込みデータを前記他のストレージ制御ユニットに送信する前記ステップは、

前記ストレージ制御ユニットが、前記他のストレージ制御ユニットが備える前記通信バッファに前記書き込み要求及び前記書き込みデータを書き込むステップと、

前記他のストレージ制御ユニットが、前記他のストレージ制御ユニットの前記通信バッファから前記書き込み要求及び前記書き込みデータを読み出すステップと

であることを特徴とする請求項７に記載の記憶デバイス制御装置の制御方法。

【請求項１１】　請求項８に記載の記憶デバイス制御装置の制御方法において、

前記ストレージ制御ユニットは、前記他のストレージ制御ユニットとの間で授受されるデータを記憶するための通信バッファを備えており、

前記ストレージ制御ユニットが、前記他のストレージ制御ユニットから前記書き込みデータの前記入出力制御を完了した旨のデータを受信する前記ステップは、

前記他のストレージ制御ユニットにより前記ストレージ制御ユニットが備える前記通信バッファに書き込まれた前記データを、前記ストレージ制御ユニットが読み出すステップであること

を特徴とする記憶デバイス制御装置の制御方法。

【請求項１２】　前記装着部には前記ストレージ制御ユニットと前記ディスク制御ユニットとが少なくとも装着されている請求項５に記載の記憶デバイス制御装置の制御方法であって、

前記ストレージ制御ユニットが、前記データ入出力要求を受信するステップと、

前記ストレージ制御ユニットが前記情報を参照し、前記入出力要求の対象となっている前記記憶ボリウムに対する前記入出力制御を行う前記ユニットを特定するステップと、

前記入出力制御を行う前記ユニットが前記ストレージ制御ユニットである場合には、前記ストレージ制御ユニットが前記入出力制御を行い、前記入出力制御を行う前記ユニットが前記ストレージ制御ユニットでない場合には、前記ディスク制御ユニットが前記入出力制御を行うステップと

を備えることを特徴とする記憶デバイス制御装置の制御方法。

【請求項１３】　前記ストレージ制御ユニットが前記受信する前記データ入

出力要求が前記データの読み出し要求であって、前記入出力制御を行う前記ユニットが前記ストレージ制御ユニットでない場合には、

前記ディスク制御ユニットが前記入出力制御を行う前記ステップは、

前記ストレージ制御ユニットが、前記読み出し要求を前記ディスク制御ユニットに送信するステップと、

前記ディスク制御ユニットが、前記読み出し要求に応じて前記入出力制御を行うステップと、

前記ストレージ制御ユニットが、前記ディスク制御ユニットから前記データを受信するステップと、

前記ストレージ制御ユニットが、前記受信した前記データを情報処理装置に送信するステップと

であることを特徴とする請求項１２に記載の記憶デバイス制御装置の制御方法。

【請求項１４】　前記ストレージ制御ユニットが前記受信する前記データ入出力要求が前記データの書き込み要求であって、前記入出力制御を行う前記ユニットが前記ストレージ制御ユニットでない場合には、

前記ディスク制御ユニットが前記入出力制御を行う前記ステップは、

前記ストレージ制御ユニットが、前記書き込み要求及び書き込みデータを前記ディスク制御ユニットに送信するステップと、

前記ディスク制御ユニットが、前記書き込み要求に応じて前記書き込みデータの前記入出力制御を行うステップと

であることを特徴とする請求項１２に記載の記憶デバイス制御装置の制御方法。

【請求項１５】　請求項１２に記載の記憶デバイス制御装置の制御方法であって、

前記ストレージ制御ユニットの前記メモリは、

前記ストレージ制御ユニットが前記入出力制御を行う前記記憶ボリウムに記憶される前記データを記憶するための第１の記憶領域と、

前記ディスク制御ユニットが前記入出力制御を行う前記記憶ボリウムに記憶される前記データを記憶するための第２の記憶領域とを有し、

前記入出力制御を行う前記ユニットが前記ストレージ制御ユニットである場合

には、前記ストレージ制御ユニットが、前記第１の記憶領域に対して前記データの入出力制御を行い、

前記入出力制御を行う前記ユニットが前記ストレージ制御ユニットでない場合には、前記ストレージ制御ユニットが、前記第２の記憶領域に対して前記データの入出力制御を行うこと

を特徴とする記憶デバイス制御装置の制御方法。

【請求項１６】　前記装着部には複数の前記ストレージ制御ユニットと前記ディスク制御ユニットとが少なくとも装着されている請求項５に記載の記憶デバイス制御装置の制御方法であって、

前記ストレージ制御ユニットが、前記データ入出力要求を受信するステップと、

前記ストレージ制御ユニットが前記情報を参照し、前記入出力要求の対象となっている前記記憶ボリウムに対する前記入出力制御を行う前記ユニットを特定するステップと、

前記入出力制御を行う前記ユニットが前記ストレージ制御ユニットである場合には、前記ストレージ制御ユニットが前記入出力制御を行い、前記入出力制御を行う前記ユニットが前記ディスク制御ユニットである場合には、前記ディスク制御ユニットが前記入出力制御を行い、前記入出力制御を行う前記ユニットが他の前記ストレージ制御ユニットである場合には、前記他のストレージ制御ユニットが前記入出力制御を行うステップと

を備えることを特徴とする記憶デバイス制御装置の制御方法。

【請求項１７】　前記装着部には前記ストレージ制御ユニットと前記ディスク制御ユニットと前記キャッシュメモリユニットとが少なくとも装着されている請求項５に記載の記憶デバイス制御装置の制御方法であって、

前記ストレージ制御ユニットの前記メモリは、

前記ストレージ制御ユニットが前記入出力制御を行う前記記憶ボリウムに記憶される前記データを記憶するための第１の記憶領域と、

前記ディスク制御ユニットが前記入出力制御を行う前記記憶ボリウムに記憶される前記データを記憶するための第２の記憶領域とを有し、

前記ストレージ制御ユニットが、前記データ入出力要求を受信するステップと、

前記ストレージ制御ユニットが前記情報を参照し、前記入出力要求の対象となっている前記記憶ボリウムに対する前記入出力制御を行う前記ユニットを特定するステップと、

前記入出力制御を行う前記ユニットが前記ストレージ制御ユニットである場合には、前記ストレージ制御ユニットが、前記第１の記憶領域に対して前記データの入出力制御を行うステップと

前記データが前記第１の記憶領域に記憶されていない場合には、前記ストレージ制御ユニットは、前記キャッシュメモリユニットに対して前記データの入出力制御を行うステップと、

前記データが前記キャッシュメモリユニットに記憶されていない場合には、前記ストレージ制御ユニットが、前記記憶ボリウムに対して前記入出力制御を行うステップと

を備えることを特徴とする記憶デバイス制御装置の制御方法。

【請求項１８】　前記装着部には前記ストレージ制御ユニットと前記ディスク制御ユニットと前記キャッシュメモリユニットとが少なくとも装着されている請求項５に記載の記憶デバイス制御装置の制御方法であって、

前記ストレージ制御ユニットの前記メモリは、

前記ストレージ制御ユニットが前記入出力制御を行う前記記憶ボリウムに記憶される前記データを記憶するための第１の記憶領域と、

前記ディスク制御ユニットが前記入出力制御を行う前記記憶ボリウムに記憶される前記データを記憶するための第２の記憶領域とを有し、

前記ストレージ制御ユニットが、前記データ入出力要求を受信するステップと、

前記ストレージ制御ユニットが前記情報を参照し、前記入出力要求の対象となっている前記記憶ボリウムに対する前記入出力制御を行う前記ユニットを特定するステップと、

前記入出力制御を行う前記ユニットが前記ディスク制御ユニットである場合に

は、前記ストレージ制御ユニットが、前記第２の記憶領域に対して前記データの入出力制御を行うステップと

前記データが前記第２の記憶領域に記憶されていない場合には、前記ストレージ制御ユニットは、前記キャッシュメモリユニットに対して前記データの入出力制御を行うステップと、

前記データが前記キャッシュメモリユニットに記憶されていない場合には、前記ディスク制御ユニットが前記記憶ボリウムに対して前記入出力制御を行うステップと

を備えることを特徴とする記憶デバイス制御装置の制御方法。

【請求項１９】　請求項５に記載の記憶デバイス制御装置の制御方法において、前記装着部に前記キャッシュメモリユニットが装着された場合には、

各前記ストレージ制御ユニットの前記メモリに記憶される、前記データ入出力要求の対象となっている前記記憶ボリウムに対する前記入出力制御を行う前記ユニットを特定するための情報の複製を、前記各ストレージ制御ユニットが、前記キャッシュメモリユニットに書き込むステップ

を備え、

前記ストレージ制御ユニットが前記情報を参照し、前記入出力要求の対象となっている前記記憶ボリウムに対する前記入出力制御を行う前記ユニットを特定する前記ステップにおいて、

前記ストレージ制御ユニットが、前記ストレージ制御ユニットの前記メモリに記憶されている前記情報を参照しても前記ユニットを特定できない場合には、前記キャッシュメモリユニットの前記情報を参照して、前記ユニットを特定すること

を特徴とする記憶デバイス制御装置の制御方法。

【請求項２０】　請求項５に記載の記憶デバイス制御装置の制御方法において、前記装着部に前記ディスク制御ユニットが装着された場合には、

ある前記ストレージ制御ユニットが、前記あるストレージ制御ユニットにより前記入出力制御が行われる前記記憶ボリウムに記憶される前記データの複製を、前記ディスク制御ユニットが前記入出力制御を行う前記記憶ボリウムに、書き込

むステップと、

前記あるストレージ制御ユニット及び前記他のストレージ制御ユニットのそれぞれが各前記メモリに記憶している、前記入出力制御を行う前記ユニットを特定するための情報を、前記あるストレージ制御ユニットから前記ディスク制御ユニットに変更するステップと

を備えることを特徴とする記憶デバイス制御装置の制御方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明は、記憶デバイス制御装置、及び記憶デバイス制御装置の制御方法に関する。

【０００２】

【従来の技術】

コンピュータシステムにおいてデータの記憶装置として用いられるストレージシステムは、ユーザのニーズに応じて小規模コンピュータシステム向けのものから大規模コンピュータシステム向けのものまで、様々なものが提供されている。

【０００３】

小規模コンピュータシステム向けのストレージシステムは、ストレージシステムとしての一通りの機能を備えた装置として提供されることにより、導入の容易化、導入時のコスト抑制が図られている。

一方、大規模コンピュータシステム向けのストレージシステムは小規模向けストレージシステムとは異なるアーキテクチャが採用され、高い拡張性を備え、最大規模での運用を求めるユーザニーズにも応えることができる構成となっている。

【０００４】

【特許文献１】

特開２００２－１２３４７９号公報

【０００５】

【発明が解決しようとする課題】

しかしながら、記憶容量の増大や、ストレージシステムの統合等によるストレージシステムの大規模化が必要な場合、小規模コンピュータシステム向けストレージシステムを採用していた場合には、大規模向けストレージシステムに入れ替えるか、ストレージシステムを追加するなどの対応が必要となる。

一方、当初から高い拡張性を備えた大規模コンピュータシステム向けストレージシステムを採用していた場合には、コンピュータシステムの導入開始初期からストレージシステムに対して大きなコストの負担を強いられる。

本発明は上記課題を鑑みてなされたものであり、記憶デバイス制御装置、及び記憶デバイス制御装置の制御方法を提供することを主たる目的とする。

【０００６】

【課題を解決するための手段】

上記課題を解決するために、本発明に係る記憶デバイス制御装置は、データ入出力要求を受信するためのホストインタフェース制御部が形成されたチャネル制御ユニットと、前記データ入出力要求に応じて、データを記憶するための記憶ボリウムに対する前記データの入出力制御を行うためのディスクインタフェース制御部が形成されたディスク制御ユニットと、前記データを記憶するためのメモリが形成されたキャッシュメモリユニットと、前記ホストインタフェース制御部と前記ディスクインタフェース制御部と前記メモリとが形成されたストレージ制御ユニットとを挿抜可能な装着部と、前記チャネル制御ユニット、前記ディスク制御ユニット、前記キャッシュメモリユニット、及び前記ストレージ制御ユニットを通信可能に接続する内部接続部とを備える。

【０００７】

本発明に係る記憶デバイス制御装置においては、ストレージ制御ユニットや、チャネル制御ユニット、ディスク制御ユニット、グローバルキャッシュのいずれも装着することができるので、顧客ニーズに応じた柔軟性の高いストレージシステム構成することができる。

【０００８】

また上記データ入出力要求とは例えばデータのリード要求やライト要求である。また入出力制御とは、データの読み出しや書き込みを行うための制御である。

記憶ボリウムとは、ハードディスク装置や半導体記憶装置等により構成される記憶装置により提供される物理的な記憶領域である物理ボリウムと、物理ボリウム上に論理的に設定される記憶領域である論理ボリウムとを含む記憶リソースである。

【０００９】

その他、本願が開示する課題、及びその解決方法は、発明の実施の形態の欄、及び図面により明らかにされる。

【００１０】

【発明の実施の形態】

以下、本発明の実施の形態について図面を用いて詳細に説明する。

＝＝＝外観構成＝＝＝

まず、本実施の形態に係るストレージシステム１００の外観構成を示す図を図１に示す。

ストレージシステム１００はディスク制御装置（記憶デバイス制御装置）１１０とディスク駆動装置１２０とを備えている。ディスク制御装置１１０はストレージシステム１００全体の制御を司る。ディスク駆動装置１２０はデータを記憶するディスクドライブ１２１を多数収納する。図１に示すストレージシステム１００ではディスク制御装置１１０が中央に配置され、その左右にディスク駆動装置１２０が配置されている。なお図１に示すように、ディスクドライブ１２１はディスク制御装置１１０にも収納されるようにすることができる。

【００１１】

ディスク制御装置１１０は、コントローラ部１１１、ファン１１３、電源部１１２を備えている。コントローラ部１１１はストレージシステム１００全体の制御を司る部分である。詳細は後述するが、コントローラ部１１１はチャネル制御ユニット３００、ディスク制御ユニット４００、ストレージ制御ユニット８００、グローバルキャッシュ（キャッシュメモリユニット）６００を含んで構成される。ディスク制御装置１１０にこれらのユニットが装着されることにより、ストレージシステム１００の制御が行われる。これらのユニットは、後述するように、一体的にユニット化された回路基板上に形成されたハードウエアさらにこのハ

ードウエアにより実行されるソフトウエア又は両方により実現される。ファン１１３はディスク制御装置１１０を冷却するために用いられる。電源部１１２はディスク制御装置１１０への電力の供給を行うために用いられる。

【００１２】

ディスク駆動装置１２０には多数のディスクドライブ１２１が収納される。ディスクドライブ１２１は、ディスク駆動装置１２０を構成する筐体に着脱可能なように収納されている。

【００１３】

また図１には示されていないが、ディスク制御装置１１０には管理端末１６０が接続されている。管理端末１６０はストレージシステム１００の保守管理を行うためのコンピュータである。管理端末１６０はストレージシステム１００に組み込まれるように構成することもできるし、遠隔地に設置されストレージシステム１００とネットワークで接続された形態とすることもできる。

【００１４】

＝＝＝全体構成＝＝＝

次に、本実施の形態に係るストレージシステムの全体構成を示すブロック図を図２に示す。

ディスク制御装置１１０はホスト計算機（情報処理装置）２００と接続され、ホスト計算機２００からのデータのリード／ライト要求（データ入出力要求）を受信する。また多数のディスクドライブ１２１と接続されており、ホスト計算機２００からのデータの入出力要求に応じて、記憶ボリウムに対してデータの入出力制御を行う。記憶ボリウムとは、記憶装置により提供される物理的な記憶領域である物理ボリウムと、物理ボリウム上に論理的に設定される記憶領域である論理ボリウム１２２とを含む記憶リソースである。記憶装置としては、例えばハードディスク装置や半導体記憶装置等様々なものを採用することができる。

【００１５】

ディスク制御装置１１０とホスト計算機２００との間の通信は、様々な通信プロトコルに従って行うようにすることができる。例えば、ファイバチャネルやＳＣＳＩ（Small Computer System Interface）、ＦＩＣＯＮ（Fibre Connection

）（登録商標）、ＥＳＣＯＮ(Enterprise System Connection)（登録商標）、ＡＣＯＮＡＲＣ（Advanced Connection Architecture）（登録商標）、ＦＩＢＡＲＣ（Fibre Connection Architecture）（登録商標）、ＴＣＰ／ＩＰ（Transmission Control Protocol/Internet Protocol）等である。これらの通信プロトコルを混在させるようにすることもできる。例えばホストＡ２００との間の通信はファイバチャネルで行い、ホストＢ２００との間の通信はＴＣＰ／ＩＰで行うようにすることもできる。ホスト計算機２００がメインフレーム計算機である場合には、例えばＦＩＣＯＮやＥＳＣＯＮ、ＡＣＯＮＡＲＣ、ＦＩＢＡＲＣが用いられる。またホスト計算機２００がオープン系の計算機である場合には、例えばファイバチャネルやＳＣＳＩ、ＴＣＰ／ＩＰが用いられる。なお、ホスト計算機２００からのデータのリード／ライト要求は、記憶ボリウムにおけるデータの管理単位であるブロックを単位として行うようにすることもできるし、ファイル名を指定することによりファイル単位に行うようにすることもできる。後者の場合にはディスク制御装置１１０は、ホスト計算機２００からのファイルレベルでのアクセスを実現するＮＡＳ（Network Attached Storage）として機能する。

【００１６】

ホスト計算機２００はＣＰＵ（Central Processing Unit）やメモリ、入出力装置等を備えたコンピュータである。ホスト計算機２００には、図示されていないクライアント計算機が接続されている。ホスト計算機２００はクライアント計算機に対して各種の情報処理サービスを提供する。ホスト計算機２００により提供される情報処理サービスは、例えば銀行の自動預金預け払いサービスやインターネットのホームページ閲覧サービスのようなオンラインサービスを始め、科学技術分野における実験シミュレーションを行うバッチ処理サービス等である。またホスト計算機２００とディスク制御装置１１０間のアクセスルートは２重化されており、片方のアクセスルートに障害が発生しても他方のアクセスルートにより入出力要求の受信を継続することができるようになっている。

【００１７】

図２に示すディスク制御装置１１０は、４枚のストレージ制御ユニット８００、２枚のチャネル制御ユニット３００、２枚のディスク制御ユニット４００、２

枚のグローバルキャッシュ６００、及び内部接続部５００を有している。またディスク制御装置１１０には管理端末１６０が接続されている。

【００１８】

＝＝＝ストレージ制御ユニット＝＝＝

ストレージ制御ユニット８００は、ホストインタフェース制御部（ホストＩＦ制御部）８１０、ディスクインタフェース制御部（ディスクＩＦ制御部）８６０、キャッシュ制御部８２０、ローカルキャッシュ（メモリ）８３０、内部インタフェース接続部（内部ＩＦ接続部）８４０を備える。ストレージ制御ユニット８００は、これらが一体的にユニット化された回路基板上に形成されたハードウエアさらにこのハードウエアにより実行されるソフトウエア又は両方により実現される。

【００１９】

ホストＩＦ制御部８１０は、ホスト計算機２００とのインタフェース機能を有する。ディスクＩＦ制御部８６０は、記憶ボリウムに対する入出力制御を行うためのインタフェース機能を備える。ローカルキャッシュ８３０は、ホスト計算機２００と記憶ボリウムとの間で授受されるデータを記憶する。キャッシュ制御部８２０は、ローカルキャッシュ８３０の制御を司る。なお本実施の形態においては、ストレージ制御ユニット８００は他のストレージ制御ユニット８００との間でクラスタを構成している。クラスタを構成することにより、同一クラスタ内のあるストレージ制御ユニット８００に障害が発生した場合でも、障害が発生したストレージ制御ユニット８００がそれまで行っていた処理を同一クラスタ内の他のストレージ制御ユニット８００に引き継ぐことにより、処理を継続することができる。キャッシュ制御部８２０は、クラスタを構成する他のストレージ制御ユニット８００のキャッシュ制御部８２０とペア間接続部８５０を介して接続されている。クラスタを構成するストレージ制御ユニット８００間でローカルキャッシュ８３０のデータを相互に記憶することによりデータの２重化を行っている。内部ＩＦ制御部８４０は、内部接続部５００を介してグローバルキャッシュ６００、ディスク制御ユニット４００、チャネル制御ユニット３００、他のストレージ制御ユニット８００に接続されている。なおストレージ制御ユニット８００は

、ホストＩＦ制御部８１０、ディスクＩＦ制御部８６０、内部ＩＦ制御部８４０を備えて構成され、ローカルキャッシュ８３０及びキャッシュ制御部８２０を備えない構成とすることもできる。この場合、クラスタを構成する各ストレージ制御ユニット８００は、例えば相互の内部ＩＦ制御部８４０をペア間接続部８５０により接続するようにすることができる。またホスト計算機２００と記憶ボリウム間で授受されるデータは、ローカルキャッシュ８３０には記憶されずに、後述するグローバルキャッシュ６００に記憶されるようにすることもできる。またローカルキャッシュ８３０、グローバルキャッシュ６００のいずれにも記憶されずに授受されるようにすることもできる。

【００２０】

本実施の形態に係るストレージ制御ユニット８００の外観構成を示す図を図３に示す。ストレージ制御ユニット８００は、ディスク制御装置１１０が備える装着部１３０に挿入されることにより、ディスク制御装置１１０に装着される。図７にストレージ制御ユニット８００がディスク制御装置１１０の装着部１３０に挿入される様子を示す。装着部１３０には複数のスロットが設けられており、各スロットにはストレージ制御ユニット８００を装着するためのガイドレールが設けられている。ガイドレールに沿ってストレージ制御ユニット８００をスロットに挿入することにより、ストレージ制御ユニット８００をディスク制御装置１１０に装着することができる。各スロットに装着されたストレージ制御ユニット８００は、ガイドレールに沿って引き抜くことにより取り外すことができる。またストレージ制御ユニット８００には、ストレージ制御ユニット８００とディスク制御装置１１０とを電気的に接続するためのコネクタ８７０が設けられている。コネクタ８７０はディスク制御装置１１０の装着部１３０の奥手方向正面部に設けられた相手側コネクタと嵌合する。

【００２１】

なおディスク制御装置１１０の各スロットには、ストレージ制御ユニット８００のみならず、チャネル制御ユニット３００、ディスク制御ユニット４００、グローバルキャッシュ６００も装着できる。いずれのユニットも、サイズやコネクタの位置、コネクタのピン配列等に互換性をもたせるようにしているからである

。従って、例えば全てのスロットにストレージ制御ユニット８００を装着するようにすることもできるし、ディスク制御ユニット４００やチャネル制御ユニット３００、グローバルキャッシュ６００を混在させて装着するようにすることもできる。

【００２２】

上述のように、ストレージ制御ユニット８００は、ホスト計算機２００とのインタフェース機能を有するホストＩＦ制御部８１０と、記憶ボリウムに対する入出力制御を行うためのインタフェース機能を備えるディスクＩＦ制御部８６０と、ホスト計算機２００と記憶ボリウムとの間で授受されるデータを記憶するローカルキャッシュ８３０とが同一パッケージ上で構成されている。これにより、ストレージ制御ユニット８００を追加していくことにより容易にシステムを拡張することが可能である。なおパッケージとは、複数の機能をモジュール化して１つの部品として構成したものである。部品の交換等の保守・管理はパッケージ単位に行われる。

【００２３】

またストレージ制御ユニット８００において、ホストＩＦ制御部８１０と、ディスクＩＦ制御部８６０と、ローカルキャッシュ８３０とが同一パッケージ上で構成されることにより、ホスト計算機２００と記憶ボリウムとの間のデータ入出力性能を向上させることができる。

【００２４】

なぜならば、これらが同一パッケージ上で構成されることにより、ホスト計算機２００と記憶ボリウムとの間のデータ転送路の電気的特性を向上させることができ、高速なデータ転送が可能となるからである。すなわち、ホストＩＦ制御部８１０、ディスクＩＦ制御部８６０、ローカルキャッシュ８３０は、ホスト計算機２００と記憶ボリウムとの間のデータ転送路の一部を構成するが、これらが同一パッケージ内に配置されることにより、データ転送路上に介在するコネクタやケーブルを削減できるため、例えばデータ転送路のインピーダンスを低下させ、耐ノイズ性を向上させることができるからである。さらに、ホストＩＦ制御部８１０、ディスクＩＦ制御部８６０、ローカルキャッシュ８３０が同一パッケージ

内で近接して配置されることにより、これらを相互に接続する配線長を短くすることができる。これによって、ストレージ制御ユニット８００内のデータ転送路のインピーダンスを低下させることができる。また耐ノイズ性も向上させることができる。これらにより、ストレージ制御ユニット８００におけるデータ転送ピッチを上げることができるようになるため、ホスト計算機２００と記憶ボリウムとの間のデータ入出力性能を向上させることができるようになる。

【００２５】

さらに、ストレージ制御ユニット８００はホスト計算機２００と記憶ボリウムとの両方に接続しているので、ホスト計算機２００からのデータ入出力要求が自己に接続している記憶ボリウムに記憶されたデータに対するものである場合には、他のユニットを介さずに処理を行うことができる。このためホスト計算機２００と記憶ボリウムとの間のデータ転送路がパッケージ間を跨るような処理が減少し、データ入出力性能を向上させることができるようになる。

【００２６】

またストレージ制御ユニット８００は、ホスト計算機２００と記憶ボリウムとの間のデータの授受をローカルキャッシュ８３０を使用せずに行うように制御することもできる。これにより、ローカルキャッシュ８３０を経由することによるデータ入出力処理の遅延を減少させることができる。これは、例えばホスト計算機２００からのデータ入出力要求に応じて行われる記憶ボリウムへのデータアクセスに局所性がない場合などローカルキャッシュ８３０を用いても高いヒット率が期待できない場合に有効である。

【００２７】

また、ストレージ制御ユニット８００を用いることにより、ホスト計算機２００と記憶ボリウムとの間でパッケージ間を跨って行われるデータ転送が減少するので、万が一ストレージ制御ユニット８００に障害が発生した場合でも、故障の影響を局所的に抑えることができる。つまりストレージ制御ユニット８００に障害が発生した場合でも、他のストレージ制御ユニット８００を用いて行われるデータ転送へ与える影響を少なくすることができる。同様に、例えば保守作業のためにストレージ制御ユニット８００の交換を行う場合にも、その影響を局所的に

抑えることができ、他のストレージ制御ユニット８００を用いて行われるデータ転送へ与える影響を少なくすることができる。

【００２８】

ストレージ制御ユニット８００を用いたディスク制御装置１１０は、初期導入時点でのコスト効果が最大になるようにしたものであり、なおかつ拡張性を維持しているため小中規模から大規模構成向きである。例えば後述するように、ストレージ制御ユニット８００を、電源１１２やファン１１３等と共に一つの筐体内に格納し、モジュラー型コントローラ１１１として構成するようにすることもできる。この場合にはストレージシステム１００の初期導入を容易に行うことができる。またシステムの拡張時もモジュラー型コントローラ１１１を順次増設することにより、容易に行うことができる。従って、これからビジネスを開始する顧客やビジネス環境の変化が激しい顧客が状況に応じてシステム規模を変更できる柔軟性のあるシステムを実現するときに有効である。またディスクＩＦ制御部８６０が記憶ボリウムから読み出したデータはキャッシュ制御部８２０を経由してローカルキャッシュ８３０のデータ領域８３１に格納される。内部接続部５００やグローバルキャッシュ６００を経由しないので、高速なデータの読み出しが可能である。

【００２９】

一方、チャネル制御ユニット３００とディスク制御ユニット４００を用いたディスク制御装置１１０は、システムの最大規模を想定して最大規模でのコスト低減効果が最大になるようにしたものであり、大規模構成向きである。従って既にビジネスが安定している顧客が比較的大規模な構成を実現するときに有効である。

【００３０】

上述のように本実施の形態に係るディスク制御装置１１０においては、ストレージ制御ユニット８００や、チャネル制御ユニット３００、ディスク制御ユニット４００、グローバルキャッシュ６００のいずれも装着することができるので、顧客ニーズに応じた柔軟性の高いストレージシステム１００を構成することができる。

【００３１】

次にストレージ制御ユニット８００の構成を示すブロック図を図８に示す。

ホストＩＦ制御部８１０は、プロセッサ８１１、メモリ８１２、ホストＩＦ回路８１４、及び内部接続ＩＦ回路８１５を備えている。プロセッサ８１１はメモリ８１２に記憶されている制御プログラム８１３を実行することにより、ホスト計算機２００とのインタフェース機能を実現する。ホストＩＦ回路８１４はホスト計算機２００と接続され、データの授受を行うための回路を構成する。内部接続ＩＦ回路８１５はキャッシュ制御部８２０との接続のための回路を構成する。

【００３２】

キャッシュ制御部８２０は、キャッシュ制御部ＩＦ回路８２１、バッファメモリ８２２、内部接続ＩＦ回路８２３、８２４、ペア間接続ＩＦ回路８２５を備える。キャッシュ制御部ＩＦ回路８２１は、ローカルキャッシュ８３０と接続するための回路を構成し、ローカルキャッシュ８３０との間のデータの授受を制御する。バッファメモリ８２２は、ローカルキャッシュ８３０との間のデータ授受の際に一時的にデータを格納するために用いられる。内部接続ＩＦ回路８２３は、ホストＩＦ制御部８１０、ディスクＩＦ制御部８６０と接続するための回路を構成する。内部接続ＩＦ回路８２４は、内部ＩＦ制御部８４０と接続するための回路を構成する。

【００３３】

ペア間接続ＩＦ回路８２５は、クラスタを構成する相手のストレージ制御ユニット８００のキャッシュ制御部８２０と接続するための回路を構成する。接続の様子を図９に示す。

クラスタを構成するストレージ制御ユニット８００は相互にローカルキャッシュ３３０のデータを共有することによりデータの２重化を行っている。データの２重化を行うためのコマンドやデータはペア間接続ＩＦ回路８２５を介して相手側のストレージ制御ユニット８００に送られる。相互のペア間接続ＩＦ回路８２５の間はペア間接続部８５０により直結されている。ペア間接続部８５０は相互のローカルキャッシュ８３０のデータを２重化するために設けられた通信路である。なおペア間接続部８５０はデータの２重化の他、例えばクラスタを構成する

ストレージ制御ユニット８００間でのメッセージ通信を行うために使用することもできる。またハートビート信号を授受するために使用するようにすることもできる。ここでハートビート信号とは、クラスタを構成する各ストレージ制御ユニット８００が相互に相手の動作状態を確認し合うための信号である。

【００３４】

ディスクＩＦ制御部８６０は、プロセッサ８６１、メモリ８６２、ディスクＩＦ回路８６４、及び内部接続ＩＦ回路８６５を備えている。プロセッサ８６１はメモリ８６２に記憶されている制御プログラム８６３を実行することにより、ディスクドライブ１２１とのインタフェース機能を実現する。ディスクＩＦ回路８６４はディスクドライブ１２１と接続され、データの授受を行うための回路を構成する。内部接続ＩＦ回路８６５はキャッシュ制御部８２０との接続のための回路を構成する。

【００３５】

なお、内部接続ＩＦ回路８１５、８６５、内部接続ＩＦ回路８２３、８２４、及びペア間接続ＩＦ回路８２５の回路構成は、同種、異種、もしくは同種・異種の混在のいずれの態様とすることもできる。

【００３６】

ローカルキャッシュ８３０は、データ領域８３１と制御領域８３２を有している。データ領域８３１は、ホスト計算機２００と記憶ボリウムとの間で授受されるデータを記憶するための記憶領域である。制御領域８３２は、データ領域８３１に記憶されているデータを管理するための記憶領域である。ローカルキャッシュ８３０の詳細については後述する。

【００３７】

＝＝＝チャネル制御ユニット＝＝＝

次にチャネル制御ユニット３００の構成を示すブロック図を図１０及び図１１に示す。またチャネル制御ユニット３００の外観構成を示す図を図４に示す。

チャネル制御ユニット３００は、ホストインタフェース制御部（ホストＩＦ制御部）３１０、キャッシュ制御部３２０、ローカルキャッシュ（メモリ）３３０、内部インタフェース接続部（内部ＩＦ接続部）３４０を備える。チャネル制御

ユニット３００は、これらが一体的にユニット化された回路基板上に形成されたハードウエアさらにこのハードウエアにより実行されるソフトウエア又は両方により実現される。

【００３８】

ホストＩＦ制御部３１０は、ホスト計算機２００とのインタフェース機能を有する。ホストＩＦ制御部３１０は、プロセッサ３１１、メモリ３１２、ホストＩＦ回路３１４、及び内部接続ＩＦ回路３１５を備えている。ホストＩＦ制御部３１０により実現される機能、及び構成等はホストＩＦ制御部８１０と同様である。

【００３９】

キャッシュ制御部３２０、ローカルキャッシュ３３０は、ホスト計算機２００と記憶ボリウムとの間で授受されるデータを記憶する。キャッシュ制御部３２０、ローカルキャッシュ３３０により実現される機能、及び構成等もストレージ制御ユニット８００におけるキャッシュ制御部８２０、ローカルキャッシュ８３０と同様である。

内部ＩＦ制御部３４０により実現される機能、構成等についても、ストレージ制御ユニット８００における内部ＩＦ制御部８４０と同様である。

【００４０】

またチャネル制御ユニット３００は、ストレージ制御ユニット８００と同様ディスク制御装置１１０が備える装着部１３０に設けられたスロットに挿入することにより、ディスク制御装置１１０に装着される。チャネル制御ユニット３００がディスク制御装置１１０の装着部１３０に挿入される様子を図７に示す。チャネル制御ユニット３００には、チャネル制御ユニット３００とディスク制御装置１１０とを電気的に接続するためのコネクタ３７０が設けられている。コネクタ３７０はディスク制御装置１１０の装着部１３０の奥手方向正面部に設けられた相手側コネクタと嵌合する。前述したように、チャネル制御ユニット３００は他のユニットとサイズやコネクタの位置、コネクタのピン配列等に互換性をもたせるようにしている。そのため、ディスク制御装置１１０の各スロットには、ストレージ制御ユニット８００、チャネル制御ユニット３００、ディスク制御ユニッ

ト４００、グローバルキャッシュ６００を混在させて装着するようにできる。

【００４１】

＝＝＝ディスク制御ユニット＝＝＝

次に、ディスク制御ユニット４００の構成を示すブロック図を図１２に示す。またディスク制御ユニット４００の外観構成を示す図を図５に示す。

ディスク制御ユニット４００は、ディスクインタフェース制御部（ディスクＩＦ制御部）４６０、内部インタフェース接続部（内部ＩＦ接続部）４４０を備える。ディスク制御ユニット４００は、これらが一体的にユニット化された回路基板上に形成されたハードウエアさらにこのハードウエアにより実行されるソフトウエアまたは両方により実現される。

【００４２】

ディスクＩＦ制御部４６０は、ディスクドライブ１２１に対する入出力制御を行うためのインタフェース機能を有する。ディスクＩＦ制御部４６０は、プロセッサ４６１、メモリ４６２、ディスクＩＦ回路４６４、及び内部接続ＩＦ回路４６５を備えている。ディスクＩＦ制御部４６０により実現される機能、及び構成等はストレージ制御ユニット８００におけるディスクＩＦ制御部８６０と同様である。

内部ＩＦ制御部４４０により実現される機能、構成等についても、ストレージ制御ユニット８００における内部ＩＦ制御部８４０と同様である。

【００４３】

また、ディスク制御ユニット４００は、ストレージ制御ユニット８００と同様ディスク制御装置１１０が備える装着部１３０に設けられたスロットに挿入することにより、ディスク制御装置１１０に装着される。ディスク制御ユニット４００がディスク制御装置１１０の装着部１３０に挿入される様子を図７に示す。ディスク制御ユニット４００には、ディスク制御ユニット４００とディスク制御装置１１０とを電気的に接続するためのコネクタ４７０が設けられている。コネクタ４７０はディスク制御装置１１０の装着部１３０の奥手方向正面部に設けられた相手側コネクタと嵌合する。前述したように、ディスク制御ユニット４００は他のユニットとサイズやコネクタの位置、コネクタのピン配列等に互換性をもた

せるようにしている。そのため、ディスク制御装置１１０の各スロットには、ストレージ制御ユニット８００、チャネル制御ユニット３００、ディスク制御ユニット４００、グローバルキャッシュ６００を混在させて装着するようにできる。

【００４４】

＝＝＝ローカルキャッシュ＝＝＝

次に、ストレージ制御ユニット８００が備えるローカルキャッシュ８３０について図１３を用いて説明する。なおチャネル制御ユニット３００が備えるローカルキャッシュ３３０についても、その機能、構成等はストレージ制御ユニット８００が備えるローカルキャッシュ８３０と同様である。

【００４５】

ローカルキャッシュ８３０は、データ領域８３１と制御領域８３２とを有している。データ領域８３１は、ホスト計算機２００と記憶ボリウムとの間で授受されるデータを記憶するための記憶領域である。制御領域８３２は、データ領域８３１に記憶されているデータを管理するための記憶領域である。

【００４６】

データ領域８３１は、ダイレクトアクセス用データ領域８３６と、通信バッファ８３７を有している。ダイレクトアクセス用データ領域８３６はさらに、自ＳＡＶＯＬ（Storage Adapter VOLume）用領域（第１の記憶領域）８３６Ａと、他ＤＡＶＯＬ（Disk Adapter VOLume）用領域（第２の記憶領域）８３６Ｂとを有している。

【００４７】

自ＳＡＶＯＬ用領域８３６Ａは、ホスト計算機２００から受信したデータ入出力要求の対象となっている記憶ボリウムが、当該データ入出力要求を受信したストレージ制御ユニット８００に接続された記憶ボリウムに対するものである場合に、ホスト計算機２００と当該記憶ボリウムとの間で授受されるデータを記憶するための領域である。

【００４８】

他ＤＡＶＯＬ用領域８３６Ｂは、ホスト計算機２００から受信したデータ入出力要求の対象となっている記憶ボリウムが、ディスク制御ユニット４００に接続

された記憶ボリウムに対するものである場合に、ホスト計算機２００と当該記憶ボリウムとの間で授受されるデータを記憶するための領域である。他ＤＡＶＯＬ用領域８３６Ｂは、ディスク制御装置１１０にディスク制御ユニット４００が装着された場合に設けられる記憶領域である。

【００４９】

通信バッファ８３７は、ホスト計算機２００から受信したデータ入出力要求の対象となっている記憶ボリウムが、当該データ入出力要求を受信したストレージ制御ユニット８００とは別のストレージ制御ユニット８００に接続された記憶ボリウムに対するものである場合に、当該別のストレージ制御ユニット８００との間でデータ入出力要求及びデータを授受するための記憶領域である。通信バッファ８３７は、ディスク制御装置１１０に異なるクラスタに属する複数のストレージ制御ユニット８００が装着された場合に設けられる記憶領域である。

【００５０】

制御領域８３２には、キャッシュ領域管理テーブル８３３、キャッシュデータ管理テーブル８３４、ボリウム管理テーブル８３５が記憶されている。図１３に示した例では、一つのキャッシュ領域管理テーブル８３３と、２つのキャッシュデータ管理テーブル８３４Ａ、８３４Ｂと、一つのボリウム管理テーブル８３５とが記載されているが、これらのテーブルは適宜複数に分割されているようにすることもできる。

【００５１】

キャッシュ領域管理テーブル８３３は、データ領域８３１に設けられる自ＳＡＶＯＬ用領域８３６Ａ、他ＤＡＶＯＬ用領域８３６Ｂ、及び通信バッファ８３７のそれぞれの記憶領域を特定するための情報を記憶したテーブルである。記憶領域を特定するための情報は例えばローカルキャッシュのアドレス情報である。図１３の例では、データ領域８３１に設定されたアドレスのうち、”００００００００番地”から”ＡＦＦＦＦＦＦＦ番地”までが自ＳＡＶＯＬ用領域８３６Ａであり、”Ｂ００００００００番地”から”ＥＦＦＦＦＦＦＦ番地”までが他ＤＡＶＯＬ用領域８３６Ｂであり、”Ｆ００００００００番地”から”ＦＦＦＦＦＦＦＦ番地”までが通信バッファ８３７であることが示されている。キャッシュ領域管

理テーブル８３３の内容を変更することにより、上記各領域の割り当てを変更することができる。例えば、ホスト計算機２００から受信するデータ入出力要求の多くが、当該データ入出力要求を受信したストレージ制御ユニット８００に接続された記憶ボリウムに対するものである場合には、自ＳＡＶＯＬ用領域８３６Ａの割り当てを増やすようにすることができる。これによりホスト計算機２００からのデータ入出力要求に対するローカルキャッシュ８３０のキャッシュヒット率が上がることが期待できるので、ストレージシステム１００の性能向上を図ることができる。キャッシュ領域管理テーブル８３３の内容の変更は、例えばストレージシステム１００の保守管理を行うオペレータにより、管理端末１６０から行うようにすることができる。

【００５２】

キャッシュデータ管理テーブル８３４は、データ領域８３１に記憶されているデータを管理するためのテーブルである。キャッシュデータ管理テーブル８３４は、データのデータブロック毎に、”Ｖａｌｉｄ”、”Ｄｉｒｔｙ”、”Ａｄｄｒｅｓｓ”、”Ｌｏｃｋ”、”Ｏｗｎｅｒ”、”Ｐｏｉｎｔｅｒ”の欄を有する。

【００５３】

なおデータ領域８３１に記憶されるデータブロックは、どのような単位で記憶されるようにすることも可能である。ディスクドライブ１２１のブロック単位やシリンダの単位、あるいはトラックの単位等に限定されるものではない。またデータブロックのサイズは可変長とすることもできるし、固定長とすることもできる。

【００５４】

”Ｖａｌｉｄ”欄は、当該データブロックのデータが有効か否かを示す。ホスト計算機２００からデータの読み出し要求があった場合に、当該データをデータ領域８３１に見つけることができても、当該データが有効でなければキャッシュアクセスはミスヒットとなる。

”Ｄｉｒｔｙ”欄は、記憶ボリウムからローカルキャッシュ８３０に読み出されたデータがホスト計算機２００により書き換えられているか否かを示す。書き

換えられている場合は、当該データをディスクドライブ１２１へ書き戻しておく必要がある。書き換えられていなければ、当該データをディスクドライブ１２１へ書き戻す必要はない。

”Ａｄｄｒｅｓｓ”欄は、ローカルキャッシュ８３０に記憶されるデータの記憶位置を示す。

【００５５】

”Ｌｏｃｋ”欄は、クラスタを構成するストレージ制御ユニット８００のローカルキャッシュ８３０間で相互に記憶されている当該データに対する処理を禁止するか否かを示すための欄である。ローカルキャッシュ８３０は、クラスタを構成する相手のローカルキャッシュ８３０とデータ２重化用の通信路であるペア間接続部８５０で接続されており、一方のローカルキャッシュ８３０に記憶されているデータが更新された場合は、ペアを組んでいるもう一方のローカルキャッシュ８３０にも２重化して記憶される。しかし２重化を完全同時に行うことはできないため、短時間ではあるが相互のローカルキャッシュ８３０に記憶されるデータに不一致が生じる。データが不一致の間に、例えば片方のローカルキャッシュ８３０から当該データがリプレース（グローバルキャッシュ６００やディスクドライブ１２１への書き戻し）されてしまうと、誤ったデータがグローバルキャッシュ６００やディスクドライブ１２１に記憶されることも起こりうる。このような問題を発生させないために”Ｌｏｃｋ”欄が設けられ、Ｌｏｃｋが有効な間は当該データに対する更新やリプレース等の制御は禁止される。

【００５６】

”Ｏｗｎｅｒ”欄は、ペアを組むローカルキャッシュ８３０のどちらが当該データを所有しているのかを示す。ペア間では相互にデータを２重化して記憶し合っているため、どちらのデータであるのかを管理するために”Ｏｗｎｅｒ”欄が設けられている。

”Ｐｏｉｎｔｅｒ”欄は、データ領域８３１に記憶されるデータと制御領域８３２に記憶されるキャッシュデータ管理テーブル８３４の対応付けを管理するための欄である。

【００５７】

ボリウム管理テーブル８３５は、ホスト計算機２００からのデータ入出力要求の対象となっている記憶ボリウムに対する入出力制御を行うユニットを特定するための情報を記憶するためのテーブルである。ボリウム管理テーブル８３５は、”ＣＡ　Ｎｏ”欄、”ｐａｔｈ　Ｎｏ”欄、”ＤＡ　Ｎｏ”欄、”Ｖｏｌｕｍｅ　Ｎｏ”欄、”ｄｒｉｖｅ　Ｎｏ”欄、”ｃｏｎｆｉｇ”欄、”ＡｃｃｅｓｓＭｅｔｈｏｄ”欄を有する。

【００５８】

”ＣＡ　Ｎｏ”欄は、ディスク制御装置１１０に装着されているストレージ制御ユニット８００のホストＩＦ制御部８１０、またはチャネル制御ユニット３００のホストＩＦ制御部３１０に付与された識別番号を記憶するための欄である。図１３の例では、ＣＡ００及びＣＡ０１が”ＣＡ　Ｎｏ”欄に記載されている。ＣＡ００及びＣＡ０１は、図２において示されるように、相互にクラスタを組むストレージ制御ユニット８００のホストＩＦ制御部８１０である。

【００５９】

”ｐａｔｈ　Ｎｏ”欄は、ホスト計算機２００からアクセス可能な論理ボリウム１２２を特定したパスに対して付与される識別番号を記憶するための欄である。本実施の形態においては、パスはストレージ制御ユニット８００やチャネル制御ユニット３００毎に付与される。従って同じパス番号でもユニットが異なれば別のパスである。パス番号をストレージシステム１００全体でユニークな番号として付与するようにすることもできる。

【００６０】

”ＤＡ　Ｎｏ”欄は、ディスク制御装置１１０に装着されているストレージ制御ユニット８００のディスクＩＦ制御部８６０、またはディスク制御ユニット４００のディスクＩＦ制御部４６０に付与された識別番号を記憶するための欄である。図１３の例では、ＤＡ００及びＤＡ０１、ＤＡ０２及びＤＡ０３、ＤＡ０４及びＤＡ０５が”ＤＡ　Ｎｏ”欄に記載されている。ＤＡ００及びＤＡ０１は、ＣＡ００及びＣＡ０１で識別されるホストＩＦ制御部８１０を含むストレージ制御ユニット８００と同一のストレージ制御ユニット８００のディスクＩＦ制御部８６０である。ＤＡ０２及びＤＡ０３は、ＣＡ００及びＣＡ０１で識別されるホ

ストＩＦ制御部８１０を含むストレージ制御ユニット８００とは異なるストレージ制御ユニット８００のディスクＩＦ制御部８６０である。ＤＡ０４及びＤＡ０５は、ＣＡ００及びＣＡ０１で識別されるホストＩＦ制御部８１０を含むストレージ制御ユニット８００とは異なるディスク制御ユニット４００のディスクＩＦ制御部４６０である。このように、ＣＡ００、ＣＡ０１で受信したホスト計算機２００からのデータ入出力要求は、自己のストレージ制御ユニット８００のディスクＩＦ制御部８６０に接続される記憶ボリウムに対してのみならず、他のストレージ制御ユニット８００やディスク制御ユニット４００のディスクＩＦ制御部８６０、４６０に接続される記憶ボリウムに対しても行われる。

【００６１】

”Ｖｏｌｕｍｅ　Ｎｏ”欄は、”ＤＡ　Ｎｏ”欄で指定されるディスクＩＦ制御部８６０、４６０に接続されている論理ボリウム１２２を特定するための欄である。

”ｄｒｉｖｅ　Ｎｏ”欄は、”ＤＡ　Ｎｏ”欄で指定されるディスクＩＦ制御部８６０、４６０に接続されているディスクドライブ１２１を特定するための欄である。

”ｃｏｎｆｉｇ”欄は、”ｄｒｉｖｅ　Ｎｏ”欄で特定されるディスクドライブ１２１上に設定されているＲＡＩＤ（Redundant Arrays of Inexpensive Disks）の構成を記憶するための欄である。

【００６２】

”ＡｃｃｅｓｓＭｅｔｈｏｄ”欄は、ホスト計算機２００から受信したデータ入出力要求の対象となっている記憶ボリウムに対する入出力制御を行う方法を特定するための欄である。「ｄｉｒｅｃｔ」と記載されている場合は、ホスト計算機２００から受信したデータ入出力要求に指定されている当該データの記憶アドレスに基づき、当該データの入出力制御を行う。「ｍｅｓｓａｇｅ」と記載されている場合は、”ＤＡ　Ｎｏ”欄で特定されるディスクＩＦ制御部８６０、４６０を含むストレージ制御ユニット８００またはディスク制御ユニット４００に対して、ホスト計算機２００から受信したデータ入出力要求を送信する。そして当該ストレージ制御ユニット８００またはディスク制御ユニット４００により入出

力制御が行われる。

【００６３】

このようにボリウム管理テーブル８３５を用いることにより、異種ユニットが混在して装着されているディスク制御装置１１０においても、ホスト計算機２００からのデータ入出力要求に対するデータ入出力制御を行うことができる。

なお、ボリウム管理テーブル８３５にはディスクドライブ１２１の領域を指定するためのアドレス情報を記憶するための欄を設けるようにすることもできる。

【００６４】

＝＝＝グローバルキャッシュ＝＝＝

次に、グローバルキャッシュ６００の構成を示すブロック図を図１４に示す。またグローバルキャッシュ６００の外観構成を示す図を図６に示す。

グローバルキャッシュ６００は、ストレージ制御ユニット８００やチャネル制御ユニット３００、ディスク制御ユニット４００と同様ディスク制御装置１１０が備える装着部１３０に設けられたスロットに挿入することにより、ディスク制御装置１１０に装着される。グローバルキャッシュ６００がディスク制御装置１１０の装着部１３０に挿入される様子を図７に示す。グローバルキャッシュ６００には、グローバルキャッシュ６００とディスク制御装置１１０とを電気的に接続するためのコネクタ６７０が設けられている。コネクタ６７０はディスク制御装置１１０の装着部１３０の奥手方向正面部に設けられた相手側コネクタと嵌合する。前述したように、グローバルキャッシュ６００は他のユニットとサイズやコネクタの位置、コネクタのピン配列等に互換性をもたせるようにしている。そのため、ディスク制御装置１１０の各スロットには、ストレージ制御ユニット８００、チャネル制御ユニット３００、ディスク制御ユニット４００、グローバルキャッシュ６００を混在させて装着するようにできる。

【００６５】

グローバルキャッシュ６００は、データ領域６０１と制御領域６０２とを有している。データ領域６０１は、ホスト計算機２００と記憶ボリウムとの間で授受されるデータを記憶するための記憶領域である。制御領域６０２は、データ領域６０１に記憶されているデータを管理するための記憶領域である。

データ領域６０１は、ダイレクトアクセス用データ領域６０６と、通信バッファ６０７を有している。

ダイレクトアクセス用データ領域６０６は、ホスト計算機２００と当該記憶ボリウムとの間で授受されるデータを記憶するための領域である。

【００６６】

通信バッファ６０７は、ストレージ制御ユニット８００間でデータ入出力要求及びデータを授受する際に使用される記憶領域である。なおストレージ制御ユニット８００とディスク制御ユニット４００との間でデータ入出力要求及びデータを授受する際に使用するようにすることもできる。またローカルキャッシュ８３０に通信バッファ８３７が設けられる場合には、グローバルキャッシュ６００に通信バッファ６０７を設けない様にすることもできる。逆にグローバルキャッシュ６００に通信バッファ６０７が設けられる場合には、ローカルキャッシュ８３０に通信バッファ８３７を設けない様にすることもできる。

【００６７】

制御領域６０２には、キャッシュ領域管理テーブル６０３、キャッシュデータ管理テーブル６０４、ボリウム管理テーブル６０５が記憶されている。図１４に示す例では、一つのキャッシュ領域管理テーブル６０３と、一つのキャッシュデータ管理テーブル６０４と、一つのボリウム管理テーブル６０５とが記載されているが、これらのテーブルは適宜複数に分割されているようにすることもできる。

【００６８】

キャッシュ領域管理テーブル６０３は、データ領域６０１に設けられるダイレクトアクセス用データ領域６０６、及び通信バッファ６０７のそれぞれの記憶領域を特定するための情報を記憶したテーブルである。記憶領域を特定するための情報は例えばグローバルキャッシュ６００のアドレス情報である。図１４の例では、データ領域６０１に設定されたアドレスのうち、”００００００００番地”から”ＡＦＦＦＦＦＦＦ番地”までがダイレクトアクセス用データ領域６０６であり、”Ｆ００００００００番地”から”ＦＦＦＦＦＦＦＦ番地”までが通信バッファ６０７であることが示されている。キャッシュ領域管理テーブル６０３の内

容の変更は、例えばストレージシステム１００の保守管理を行うオペレータにより、管理端末１６０から行うようにすることができる。これによりホスト計算機２００からのデータ入出力要求の特性に応じたグローバルキャッシュ６００設定が行えるので、ストレージシステム１００の性能向上を図ることができる。

【００６９】

キャッシュデータ管理テーブル６０４は、データ領域６０１に記憶されているデータを管理するためのテーブルである。キャッシュデータ管理テーブル６０４の基本的な構成はローカルキャッシュ８３０のキャッシュデータ管理テーブル８３４の場合と同様であるが、”Ｌｏｃｋ”欄と”Ｏｗｎｅｒ”欄の表す意味が異なる。

【００７０】

”Ｌｏｃｋ”欄は、グローバルキャッシュ６００上の当該データがローカルキャッシュ８３０に読み出されており、ホスト計算機２００により更新される可能性があるため、他のローカルキャッシュ８３０への読み出しは禁止されている状態であることを表す。複数のローカルキャッシュ８３０にデータの読み出しを許してしまうと、それぞれ独立にホスト計算機２００によって更新される可能性があり、データの一致性が保証できなくなるためである。

”Ｏｗｎｅｒ”欄は、当該データを読み出し中のローカルキャッシュ８３０を表す。

【００７１】

グローバルキャッシュ６００は内部接続部５００に接続されており、２つのグローバルキャッシュ６００がペアとなってデータの２重化を行っている。グローバルキャッシュ６００間のデータの２重化は、内部接続部５００を介して相互にデータを転送することにより実現される。

グローバルキャッシュ６００におけるボリウム管理テーブル６０５は、ローカルキャッシュ８３０のボリウム管理テーブル８３５が複製されたものである。ローカルキャッシュ８３０が複数ある場合には、各ローカルキャッシュ８３０のボリウム管理テーブルの複製の結合である。

【００７２】

これにより、例えばあるストレージ制御ユニット８００がホスト計算機２００からデータ入出力要求を受信した場合に、当該ストレージ制御ユニット８００のローカルキャッシュ８３０を参照しても、データ入出力要求の対象となっている記憶ボリウムに対するユニットが特定できなかった場合に、グローバルキャッシュ６００のボリウム管理テーブル６０５を参照することにより記憶ボリウムに対する入出力制御を行うユニットを特定することができる。

【００７３】

＝＝＝内部接続部＝＝＝

次に、本実施の形態に係る内部接続部５００の構成を示すブロック図を図１５に示す。

内部接続部５００は、ストレージ制御ユニット８００、チャネル制御ユニット３００、ディスク制御ユニット４００、グローバルキャッシュ６００を相互に結合するためのスイッチである。

図１５には４入力４出力の場合を示したが、実際の入出力数はディスク制御装置１１０に装着可能なユニットの数に応じた数となる。

【００７４】

内部接続部５００は、受信部５１０、送信部５２０、制御部５３０を備えている。受信部５１０は、内部接続部５００に入力されてきたデータを適宜バッファ５１１に蓄えつつ、制御部５３０からの指令に従って、指定された送信部５２０のバッファ５２１へデータを転送する。送信部５２０はバッファ５２１に格納されたデータを順次出力する。図１５では内部接続部５００としてクロスバスイッチの構成をとった場合を示したが、クロスバスイッチの構成に限られることはなく、様々な構成をとることが可能である。例えば受信部５１０と送信部５２０の間を多段のスイッチ回路で接続するようにすることもできる。

【００７５】

＝＝＝管理端末＝＝＝

次に、本実施の形態に係る管理端末１６０の構成を示すブロック図を図１６に示す。

管理端末１６０は、ＣＰＵ１６１、メモリ１６２、ポート１６３、記録媒体読

取装置１６４、入力装置１６５、出力装置１６６、記憶装置１６８を備える。

【００７６】

ＣＰＵ１６１は管理端末１６０の全体の制御を司るもので、記憶装置１６８に格納された管理プログラム１６９を適宜メモリ１６２に読み出して実行することによりストレージシステム１００の保守管理のための各種機能を実現する。例えばディスクドライブ１２１上への論理ボリウム１２２の設定や、ストレージ制御ユニット８００のホストＩＦ制御部８１０において実行されるプログラム８１３のインストール等を行うことができる。記録媒体読取装置１６４は、記録媒体１６７に記録されているプログラムやデータを読み取るための装置である。読み取られたプログラムやデータはメモリ１６２や記憶装置１６８に格納される。従って、例えば記録媒体１６７に記録された管理プログラム１６９やプログラム８１３等を、記録媒体読取装置１６４を用いて上記記録媒体１６７から読み取って、メモリ１６２や記憶装置１６８に格納するようにすることができる。記録媒体１６７としてはフレキシブルディスクやＣＤ－ＲＯＭ、ＤＶＤ－ＲＯＭ、半導体メモリ等を用いることができる。記録媒体読取装置１６４は管理端末１６０に内蔵されている形態とすることもできるし、外付されている形態とすることもできる。記憶装置１６８には管理プログラム１６９が記憶されている。記憶装置１６８は、例えばハードディスク装置や半導体記憶装置等である。入力装置１６５はオペレータ等による管理端末１６０へのデータ入力等のために用いられる。入力装置１６５としては例えばキーボードやマウス等が用いられる。出力装置１６６は情報を外部に出力するための装置である。出力装置１６６としては例えばディスプレイやプリンタ等が用いられる。ポート１６３はディスク制御装置１１０と通信を行うための装置である。また図示されていない他のコンピュータとの間で通信を行うために使用することもできる。この場合、例えばプログラム８１３をポート１６３を介して他のコンピュータから受信して、ストレージ制御ユニット８００にインストールするようにすることもできる。

【００７７】

＝＝＝コントローラの増設＝＝＝

上述したように、本実施の形態に係るストレージシステム１００においては、

ストレージ制御ユニット８００、チャネル制御ユニット３００、ディスク制御ユニット４００、グローバルキャッシュ６００をディスク制御装置１１０に混在させて装着させることが可能である。これにより、顧客毎に異なるストレージシステム１００の構成に対する要求に柔軟に対応することが可能となる。例えば、ストレージシステム１００の導入時には少数のディスクドライブ１２１とストレージ制御ユニット８００等により小規模のストレージシステム１００を構成し、顧客の事業の拡大等の際にストレージ制御ユニット８００や、チャネル制御ユニット３００、ディスク制御ユニット４００、グローバルキャッシュ６００を追加するようにする。これにより顧客の要望に応じてストレージシステム１００の規模拡大を行うことができる。その様子を図１７に示す。またストレージシステム１００の規模拡大前の導入時のシステム構成図を図１８に示す。

【００７８】

図１８に示す例では、導入時のディスク制御装置１１０は、クラスタを組むストレージ制御ユニット８００と、内部接続部５００と、管理端末１６０とにより構成されている。図１８に示すディスク制御装置１１０において、導入時から高価な内部接続部５００が用いられているのは、後から内部接続部５００を追加保守するのは困難であるほか、ディスク制御装置１１０を大掛かりに分解した後の再組み立てが必要となり事実上不可能なためである。しかし図１９に示すように導入時には内部接続部５００を設けないように構成するようにすることも可能である。しかしながらこの場合、例えばグローバルキャッシュ６００を増設する場合には、図２０に示すように内部接続部５００の増設も必要となる。

【００７９】

また、ストレージシステム１００の導入をより容易に行うことができるように、初期コントローラ１１１を採用するようにすることもできる。その様子を図２１乃至図２３に示す。

【００８０】

図２１に示すように、導入時には初期コントローラ１１１と少数のディスクドライブ１２１で運用を開始し、その後のシステム拡大時に装着部１３０にストレージ制御ユニット８００やディスク制御ユニット４００等を増設するようにする

こともできる。初期コントローラ１１１は、図２２に示すように一つの筐体のなかにストレージ制御ユニット（ＳＡ）８００や電源１１２、ファン１１３が収納されたものであり、モジュラー型コントローラ１１１として提供されるものである。これにより、初期コントローラ１１１と必要な容量のディスクドライブ１２１を導入時に用意することにより、ストレージシステム１００の運用を開始することができる。

【００８１】

図２１、図２２に示す初期コントローラ１１１を用いた場合のストレージシステム１００のシステム構成を図２３に示す。図２３に示すようにこの場合は、ストレージ制御ユニット８００はモジュラー型コントローラ１１１すなわち初期コントローラ１１１として提供される。この場合ストレージシステム１００の最大構成を想定したコストの高い内部接続部５００を併せ持つ必要はない。すなわち一体型コントローラ１１１を導入する必要は無い。このため低コストでストレージシステム１００を実現できる効果がある。またグローバルキャッシュ６００やディスク制御ユニット４００は装着部１３０に挿入される一体型コントローラとして提供される。この場合内部接続部５００と初期コントローラ１１１間は、ケーブルにより接続されることになる。

【００８２】

また図２４に示すように、ストレージシステム１００の導入時に、初期コントローラ１１１の他に、モジュール化されたＳＷ（Switch）を装着するようにすることができる。ＳＷは内部接続部５００を構成する装置である。この場合のシステム構成図を図２５に示す。図２５に示すように、この場合の内部接続部５００は、破線で囲んだように、モジュール化されて装着されたＳＷと、一体型コントローラ１１１の導入時に追加されたＳＷとの組み合わせにより構成される。そして上記各ＳＷ間やＳＷとストレージ制御ユニット８００間はケーブルにより接続されることになる。

【００８３】

＝＝＝データ入出力処理の流れ＝＝＝

次に、本実施の形態に係るストレージシステム１００がホスト計算機２００か

らデータ入出力要求を受信した場合に行われるデータ入出力処理の流れについて説明する。なお本実施の形態に係るデータ入出力処理は、各種の動作を行うためのコードから構成される制御プログラム８１３、８６３、３１３、４６３を、それぞれプロセッサ８１１、８６１、３１１、４６１が実行することにより実現される。

【００８４】

まず本実施の形態に係るストレージシステム１００において、グローバルキャッシュ６００が増設された場合に行われる、ストレージ制御ユニット８００、チャネル制御ユニット３００のローカルキャッシュ８３０、３３０のボリウム管理テーブル８３５、３３５のグローバルキャッシュ６００への移行処理の流れについて、図２６を参照しながら説明する。この処理は、管理端末１６０からの指示により、ストレージ制御ユニット８００やチャネル制御ユニット３００のプロセッサ８１１、３１１により行われる。

【００８５】

まずグローバルキャッシュ６００にボリウム管理テーブル６０５を作成するための領域を確保する（Ｓ１０００）。次にローカルキャッシュ８３０、３３０のボリウム管理テーブル８３５、３３５をロックする（Ｓ１００１）。そしてローカルキャッシュ８３０、３３０のボリウム管理テーブル８３５、３３５の複製をグローバルキャッシュ６００に書き込む（Ｓ１００２）。書き込みが終了したらローカルキャッシュ８３０、３３０のボリウム管理テーブル８３５，３３５のロックを解放して、処理を終了する（Ｓ１００３）。なおＳ１０００において確保する領域は、少なくともコピーするボリウム管理テーブル８３５、３３５よりも大きくする。またロックする単位はコピー単位とすることもできる。

【００８６】

これにより、例えばあるストレージ制御ユニット８００がホスト計算機２００からデータ入出力要求を受信した場合に、当該ストレージ制御ユニット８００のローカルキャッシュ８３０を参照しても、データ入出力要求の対象となっている記憶ボリウムに対するユニットが特定できなかった場合に、グローバルキャッシュ６００のボリウム管理テーブル６０５を参照することにより記憶ボリウムに対

する入出力制御を行うユニットを特定することができる。

【００８７】

次に新規に記憶ボリウムが設定された場合に行われる、ボリウム管理テーブル８３５、３３５の更新処理について図２７を参照しながら説明する。

まず、新規に設定された記憶ボリウムに対する入出力制御を行うユニットの種類が、ストレージ制御ユニット８００であるかどうかの判定を行う（Ｓ２０００）。ストレージ制御ユニット８００である場合には、ローカルキャッシュ８３０のボリウム管理テーブル８３５をロックする（Ｓ２００１）。そして新規に追加された記憶ボリウムに関する情報をボリウム管理テーブル８３５に書き込む（Ｓ２００２）。そして書き込みが終了したらローカルキャッシュ８３０のボリウム管理テーブル８３５のロックを解放する（Ｓ２００３）。続いてグローバルキャッシュ６００のボリウム管理テーブル６０５に対してもローカルキャッシュ８３０のボリウム管理テーブル８３５と同様の書き込みを行って処理を終了する（Ｓ２００４乃至Ｓ２００６）。なおＳ２０００において、ストレージ制御ユニット８００でない場合には、グローバルキャッシュ６００のボリウム管理テーブル６０５に対してのみ処理を行う。

【００８８】

次に、ホスト計算機２００から本実施例に係るストレージシステム１００に対してデータアクセス要求があった場合の処理の流れを示すフローチャートを図２８に示す。

ホスト計算機２００からストレージ制御ユニット８００またはチャネル制御ユニット３００にデータのアクセス要求があると（ステップＳ３０００）、ホストＩＦ制御部８１０、３１０のプロセッサ８１１、３１１はアクセス要求の解析を行う。解析によりアクセスの種類（リード要求、ライト要求）やアクセスするデータのアドレス等を判別する。

【００８９】

続いてプロセッサ８１１、３１１は、アクセスの種類に応じて、図３８に示すコマンドをキャッシュ制御部８２０、３２０のキャッシュ制御部ＩＦ回路８２１、３２１に送信する。リード要求の場合は図３８Ｄのコマンドが送信され、ライ

ト要求の場合は図３８Ａのコマンドと図３８Ｂのコマンド（データ）が送信される。なお、図３８に示すコマンドのフォーマットを図３７に示す。図３７に示すように図３８に示すコマンドはヘッダー部とペイロード部を備えている。ヘッダ部は転送先アドレス、転送元アドレス、転送長、パケット種別を備えている。パケットの種別は、図３８に示すようにＷＲＩＴＥ、ＲＥＡＤ、データ、ステータス情報である。

【００９０】

キャッシュ制御部ＩＦ回路８２１、３２１はプロセッサ８１１、３１１から送信されたコマンドに従い、ローカルキャッシュ８３０、３３０の制御領域８３２、３３２に記録されているボリウム管理テーブル８３５を参照して、データ入出力要求の対象となっている記憶ボリウムに対する入出力制御を行うユニットを特定する。そしてボリウム管理テーブル８３５の”ＡｃｃｅｓｓＭｅｔｈｏｄ”欄を参照して、当該ユニットに対するデータアクセス方法を特定する（Ｓ３００１）。「ｄｉｒｅｃｔ」と記載されていた場合には、Ｓ３００２に進む。

【００９１】

Ｓ３００２において、キャッシュ制御部ＩＦ回路８２１、３２１はプロセッサ８１１、３１１から送信されたコマンドに従い、ローカルキャッシュ８３０、３３０の制御領域８２２、３３２に記録されているキャッシュデータ管理テーブル８３４、３３４を検索して、コマンドに指定されたアドレスのデータがローカルキャッシュ８３０、３３０に記憶されているかどうかを確認する（ステップＳ３００２）。

【００９２】

当該データがローカルキャッシュ８３０、３３０にある（ヒット）場合は（Ｓ３００２）、ローカルキャッシュ８３０、３３０に対するリードライト処理を行い（Ｓ３００７）、ホスト計算機２００へ完了報告を行う（Ｓ３００８）。

【００９３】

Ｓ３００２において行われる処理を図２９に示す。まずＳ４０００においてホスト計算機２００からのデータ入出力要求の対象となっている記憶ボリウムに対する入出力制御を行うユニットを判定する（Ｓ４０００）。自ユニット配下の記

憶ボリウムに対するものである場合には、自ボリウム用領域（第１の記憶領域）を対象にヒットミスの判定を行う（Ｓ４００１）。ミスヒットの場合はグローバルキャッシュ６００または記憶ボリウムから当該データをステージングする（Ｓ４００２、Ｓ４００３）。ステージングとは下位階層の記憶装置からデータを読み出してくることをいう。一方Ｓ４０００において、他ユニット配下の記憶ボリウムに対するものである場合には、他ボリウム用領域（第２の記憶領域）を対象にヒットミスの判定を行う（Ｓ４００４）。ミスヒットの場合はグローバルキャッシュ６００または記憶ボリウムから当該データをステージングする（Ｓ４００５、Ｓ３００３）。

【００９４】

またＳ３００７において行われるローカルキャッシュ８３０、３３０に対するリードライト処理の流れについて図３３を参照しながら説明する。

ホスト計算機２００からのアクセス要求がリード要求の場合であれば、キャッシュ制御部ＩＦ回路８２１、３２１は当該データをローカルキャッシュ８３０、３３０から読み出してホスト計算機２００へ送信する（Ｓ８０００、Ｓ８００１）。ローカルキャッシュ８３０、３３０から読み出し完了の報告（ＡＣＫ）を受けると、キャッシュ制御部ＩＦ回路８２１、３２１はプロセッサ８１１、３１１に対してステータスを送信する。送信されるステータスは図３８Ｆで示されるコマンドである。最後にプロセッサ８１１、３１１はホスト計算機２００にデータの読み出し完了報告を行って（Ｓ３００８）処理を終了する。以上のリード要求の処理をフローチャートで表したものを図４０に示す。

【００９５】

一方、ホスト計算機２００からのアクセス要求がライト要求の場合は、キャッシュ制御部ＩＦ回路８２１、３２１は、ホスト計算機２００から送信されバッファメモリ８２２、３２２に格納されている書き込みデータをローカルキャッシュ８３０、３３０に書き込む（Ｓ８００２）。ローカルキャッシュ８３０、３３０への書き込み処理の詳細は図３４に示される。すなわち、まずキャッシュ制御部ＩＦ回路８２１、３２１は、ペアとなっている相手側のキャッシュ制御部ＩＦ回路８２１、３２１に対してローカルキャッシュ８３０、３３０をロックするよう

に要求を出す。相手からロック確保の応答を受け取り、自他ローカルキャッシュ３３０のロックを確保すると（Ｓ９０００）、キャッシュ制御部ＩＦ回路８２１、３２１は、バッファメモリ８２２、３２２に格納されている書き込みデータを、ペア間接続部８５０、３５０を介して相手側のバッファメモリ８２２、３２２に送信する。そして相手側のキャッシュ制御部ＩＦ回路８２１、３２１により相手側のローカルキャッシュ８３０、３３０に書き込みが行われる（Ｓ９００１）。続いて自分側のローカルキャッシュ８３０、３３０にデータの書き込みを行う（Ｓ９００２）。なお、データをローカルキャッシュ８３０、３３０に書き込む際にはキャッシュデータ管理テーブル８３４、３３４の”Ｄｉｒｔｙ”欄にチェックを入れる。相互のローカルキャッシュ８３０、３３０にデータの書き込みが完了すると、ロックを解除した後、ホスト計算機２００へ完了報告を送信し、処理を終了する（Ｓ９００３）。以上のライト要求の処理をフローチャートで表したものを図４０に示す。

【００９６】

なお本実施例においては、ローカルキャッシュ８３０、３３０のキャッシュデータ管理テーブル８３４、３３４の検索や、ローカルキャッシュ８３０、３３０からのデータの読み出し等の制御は、キャッシュ制御部ＩＦ回路８２１、３２１が行う場合を例に説明したが、プロセッサ８１１、３１１が行う態様とすることも可能である。

【００９７】

また詳細は後述するが、グローバルキャッシュ６００へのデータアクセス制御についても、キャッシュ制御部ＩＦ回路８２１、３２１が行う態様に限らず、プロセッサ８１１、３１１が行う態様とすることも可能である。

【００９８】

次に、ホスト計算機２００からのデータ入出力要求を受けたが、ローカルキャッシュ８３０、３３０に当該データが無い場合、すなわちキャッシュミスヒットの場合の処理について説明する。

この場合はグローバルキャッシュ６００に当該データがあるかどうかを確認する（Ｓ３００３）。まずキャッシュ制御部ＩＦ回路８２１、３２１はプロセッサ

８１１、３１１から送信されたコマンドに指定された当該データのアドレスを元に、内部接続部５００を介してグローバルキャッシュ６００にコマンドを送信する。そしてグローバルキャッシュ６００の制御領域６０２に記録されているキャッシュデータ管理テーブル６０４検索して、当該データがグローバルキャッシュ６００に記憶されているかどうかを確認する。

【００９９】

当該データがグローバルキャッシュ６００に無ければ、ボリウム管理テーブル６０５を参照して、データ入出力要求の対象となっている記憶ボリウムに対する入出力制御を行うユニットに対してコマンドを送信する。そして当該データを記憶ボリウムから読み出して、グローバルキャッシュ６００に格納させる（Ｓ３００４）。グローバルキャッシュ６００に格納されたデータは、内部接続部５００を介してもう１つのグローバルキャッシュ６００へも送られ、データの２重化が行われる。

【０１００】

なおここで、記憶ボリウムからグローバルキャッシュ６００へ読み出したデータをいち早くホスト計算機２００へ届けるための処理を優先させ、グローバルキャッシュ６００上でのデータの２重化を後回しにする態様とすることも可能である。かかるグローバルキャッシュ６００上のデータは記憶ボリウムにも記憶されているので、グローバルキャッシュ６００上で消失しても問題は生じないからである。当該データが更新された場合に２重化を行うようにすることでデータの信頼性は確保できる。

【０１０１】

続いて当該データをグローバルキャッシュ６００上でロックを掛ける（Ｓ３００５）。すなわち、グローバルキャッシュ６００上の当該データを他のローカルキャッシュ８３０、３３０から読み出されないようにする。その処理の流れを図３０のフローチャートに示す。

当該データがすでに他のローカルキャッシュ８３０、３３０に読み出されており、ロックが掛けられている場合には（Ｓ５０００）、当該ローカルキャッシュ８３０、３３０に対してロックを解放するように要求する（Ｓ５００１）。どの

ローカルキャッシュ８３０、３３０がロックを掛けているのかは、キャッシュデータ管理テーブル６０４の”Ｏｗｎｅｒ”欄で知ることができる。ロックが解放されるのを待った後（Ｓ５００２）、ロックを掛け、他のローカルキャッシュ８３０、３３０から読み出されないようにしてから処理を終了する（Ｓ５００３）。他のローカルキャッシュ８３０、３３０にロックが掛けられていなければ、直ちにロックを掛けて処理を終了する（Ｓ５０００、Ｓ５００３）。

【０１０２】

続いてグローバルキャッシュ６００上に格納された当該データをローカルキャッシュ８３０、３３０へ読み出す処理を行う（Ｓ３００６）。その処理の流れを図３１のフローチャートに示す。

まずグローバルキャッシュ６００からローカルキャッシュ８３０、３３０にデータを転送する前に、ローカルキャッシュ８３０、３３０上に当該データを書き込むためのキューの空きスロットがあるかどうかを調べる（Ｓ６０００）。ここでスロットとはキューを構成する個々の記憶領域を言う。なおこの処理は、ローカルキャッシュ８３０、３３０上に当該データを書き込むための空き領域があるかどうかを調べるようにすることもできる。この場合は、キャッシュデータ管理テーブル８３４、３３４の”Ｖａｌｉｄ”欄を検索して無効なデータの総容量がグローバルキャッシュ６００から転送されるデータの総容量よりも大きいかどうかをチェックする。

【０１０３】

空きスロットがある場合には、まずキャッシュ制御部ＩＦ回路８２１、３２１は、ペアとなっている相手側のキャッシュ制御部ＩＦ回路８２１、３２１に対してローカルキャッシュ８３０、３３０をロックするように要求を出してロックを確保する（Ｓ６００２）。次にグローバルキャッシュ６００からデータをバッファメモリ８２２、３２２に格納し、ペア間接続部８５０、３５０を介して相手側のバッファメモリ８２２、３２２にデータを送信するとともに、自分のローカルキャッシュ８３０、３３０にもデータの書き込みを行う（Ｓ６００３、Ｓ６００４）。相互のローカルキャッシュ８３０、３３０にデータの書き込みが完了すると、ロックを解除して処理を終了する（Ｓ６００５）。この後の処理はホスト計

算機２００からのデータ入出力要求に応じて、上述した通りに行われる（Ｓ３００７、Ｓ３００８）。

【０１０４】

なお、グローバルキャッシュ６００からローカルキャッシュ８３０、３３０にデータを転送するための空きスロットがない場合は、ローカルキャッシュ８３０、３３０上のいずれかのデータをグローバルキャッシュ６００に書き戻すことにより空きスロットを確保する処理が必要になる（Ｓ６００１）。その処理の流れを図３２のフローチャートに示す。

まずキャッシュ制御部ＩＦ回路８２１、３２１は、ペアとなっている相手側のキャッシュ制御部ＩＦ回路８２１、３２１に対してローカルキャッシュ８３０、３３０をロックするように要求を出してロックを確保する（Ｓ７０００）。続いて所定のアルゴリズムにより特定したグローバルキャッシュ６００へ書き出されるデータのＤｉｒｔｙビットをキャッシュデータ管理テーブル８３４、３３４により調べる（Ｓ７００１）。所定のアルゴリズムとしては、最も長期間アクセスのなかったデータをキャッシュから書き出すＬＲＵ（Least Recently Used）方式が一般的であるが、他のアルゴリズムとすることもできる。

【０１０５】

Ｄｉｒｔｙビットがセットされていなければグローバルキャッシュ６００へデータを書き出す必要はないが、Ｄｉｒｔｙビットがセットされている場合はグローバルキャッシュ６００へデータを書き出す必要があるので、グローバルキャッシュ６００に当該データを書き込むための空きスロットがあるかどうかを調べる（Ｓ７００２）。グローバルキャッシュ６００に空きスロットがない場合はグローバルキャッシュ６００のデータを記憶ボリウムへ書き出して、空きスロットを確保する（Ｓ７００３）。

【０１０６】

続いてグローバルキャッシュ６００上の空きスロットにローカルキャッシュ８３０、３３０からデータを書き出す（Ｓ７００４）。書き出しは２つのグローバルキャッシュ６００に対して行われる。グローバルキャッシュ６００にデータが書き出された後は、もはや当該データはＤｉｒｔｙではないので、Ｄｉｒｔｙビ

ットをリセットする（Ｓ７００５）。続いて当該データが記憶されていたローカルキャッシュ８３０、３３０上のスロットを解放する必要がある場合には（Ｓ７００６）、当該データのＶａｌｉｄビットをリセットする（Ｓ７００７）。

【０１０７】

そしてペアとなっている相手のローカルキャッシュ８３０、３３０に対して当該データのグローバルキャッシュ６００への書き出しが完了した旨の報告を行う（Ｓ７００８）。この報告をうけた相手側のローカルキャッシュ８３０、３３０では、キャッシュデータ管理テーブル８３４、３３４のＶａｌｉｄビットがリセットされる。最後にローカルキャッシュ８３０、３３０のロックを解放して（Ｓ７００９）処理を終了する。

【０１０８】

一方、Ｓ３００１においてボリウム管理テーブル８２５、３２５の”ＡｃｃｅｓｓＭｅｔｈｏｄ”欄に「ｍｅｓｓａｇｅ」と記載されていた場合には、Ｓ３００９に進む。Ｓ３００９の処理の流れを図３５に示す。

【０１０９】

まずキャッシュ制御部ＩＦ回路８２１、３２１は、メッセージアクセス用の領域（通信バッファ）を確保する（Ｓ１００００）。すなわち、データ入出力要求の対象となっている記憶ボリウムに対する入出力制御を行うユニットのローカルキャッシュ８３０、３３０の通信バッファ８３７の空き領域を確保する。

【０１１０】

ホスト計算機２００からのデータ入出力要求がリード要求の場合には、上記領域を確保した通信バッファ８３７に当該データ入出力要求を書きこむ（Ｓ１０００１、Ｓ１００００３）。データ入出力要求の書きこみは、図３９に示すメッセージにより行われる。図３９Ａに示すメッセージにより、メッセージコマンドであることが示される。そして図３９Ｂに示すメッセージのメッセージデータ欄に当該データ入出力要求が挿入されて、上記通信バッファ８３７に書きこまれる。データ入出力制御の終了通知及び読み出されたデータを受け取ったら（Ｓ１０００４）、ホスト計算機２００に当該データを送信する（Ｓ１００００６）。

【０１１１】

一方、ホスト計算機２００からのデータ入出力要求がライト要求の場合には、上記領域を確保した通信バッファ８３７に当該データ入出力要求及び書き込みデータを書きこむ（Ｓ１０００１乃至Ｓ１０００３）。そして通信バッファ８３７にデータ入出力制御の終了通知が書き込まれたら（Ｓ１０００４）、ホスト計算機２００に当該終了通知を送信して処理を終了する。

【０１１２】

自己の通信バッファ８３７にメッセージを書き込まれたストレージ制御ユニット８００により行われる処理を図３６に示す。

まず、キャッシュ制御部ＩＦ回路８２１、３２１は、自己の通信バッファ８３７にメッセージを書き込まれたことを検知すると（Ｓ１１００１）、通信バッファ８３７からデータ入出力要求を読み出す（Ｓ１１００２）。次に当該データ入出力要求の対象となっているデータがローカルキャッシュ８３０に記憶されているか否かをチェックする（Ｓ１１００３）。ミスヒットの場合は、グローバルキャッシュ６００または記憶ボリウムから当該データを読み出してきてローカルキャッシュ８３０に記憶する（Ｓ１１００４）。そしてデータ入出力要求がリード要求の場合は、当該読み出したデータをメッセージ送信元の通信バッファ８３７に書き込む（Ｓ１１００６）。またデータ入出力要求がライト要求の場合は、当該データ入出力要求に従ってデータをローカルキャッシュ８３０へ書き込む（Ｓ１１００７）。この処理は図３４に示した処理と同様である。そして書き込み完了通知を相手側の通信バッファ８３７に書き込む。

【０１１３】

なお、通信バッファ８３７を介してストレージ制御ユニット８００間でメッセージの授受を行う際の処理の流れを図４１に示す。

通信バッファ８３７を介してデータ入出力要求の授受を行うようにすることにより、各ストレージ制御ユニット８００は、他のストレージ制御ユニット８００の処理に依存せずに、当該他のストレージ制御ユニットに接続されている記憶ボリウムのデータに対するデータ入出力制御を行うことが可能となる。

【０１１４】

次に、ストレージシステム１００の導入時にストレージ制御ユニット８００に

より運用を行っていたストレージシステム１００において、ディスク制御ユニット４００が増設された場合に行うことができる、記憶ボリウムの変更処理について図４２を参照しながら説明する。

【０１１５】

この処理は、ストレージ制御ユニット８００に接続されている記憶ボリウムに記憶されているデータの複製を、ディスク制御ユニット４００に接続されている記憶ボリウムに書き込むことにより、それ以降のホスト計算機２００からのデータ入出力要求に対する入出力制御を当該ディスク制御ユニット４００に接続された記憶ボリウムに対して行うようにするものである。このようにすることにより、ストレージシステム１００の構成変更の柔軟性を高めることができるようになる。例えば、ストレージシステム１００の導入時には初期コントローラ１１１により運用していた顧客が、その後のストレージシステム１００の規模拡大時に、チャネル制御ユニット３００とディスク制御ユニット４００とを用いた、拡張性の高いシステム構成に変更するようにすることが出来るようになる。

【０１１６】

この処理は、管理端末１６０からの指示により、ストレージ制御ユニット８００やチャネル制御ユニット３００のプロセッサ８１１により行われる。

まずプロセッサ８１１は、ローカルキャッシュ８３０のボリウム管理テーブル８３５、及びグローバルキャッシュ６００のボリウム管理テーブル６０５をロックする（Ｓ１６０００、Ｓ１６００１）。そして、ローカルキャッシュ８３０のボリウム管理テーブル８３５の”ＤＡ　Ｎｏ”欄、”ｖｏｌｕｍｅ　Ｎｏ”欄、”ｄｒｉｖｅ　Ｎｏ”欄を、ストレージ制御ユニット８００に接続されている記憶ボリウムに関する情報から、ディスク制御ユニット４００に接続されている記憶ボリウムに関する情報に変更する（Ｓ１６００２）。またグローバルキャッシュ６００のボリウム管理テーブル６０５の”ＤＡ　Ｎｏ”欄、”ｖｏｌｕｍｅ　Ｎｏ”欄、”ｄｒｉｖｅ　Ｎｏ”欄を、ストレージ制御ユニット８００に接続されている記憶ボリウムに関する情報から、ディスク制御ユニット４００に接続されている記憶ボリウムに関する情報に変更する（Ｓ１６００３）。そしてローカルキャッシュ８３０のボリウム管理テーブル８３５、及びグローバルキャッシュ

６００のボリウム管理テーブル６０５をロックを解放する（Ｓ１６００４、Ｓ１６００５）。

【０１１７】

これにより、ホスト計算機２００からのデータ入出力要求に対する入出力制御を、ストレージ制御ユニット８００に接続された記憶ボリウムから、ディスク制御ユニット４００に接続された記憶ボリウムに対して行うようにすることができるようになる。

【０１１８】

以上のように本実施の形態に係るディスク制御装置１１０においては、ストレージ制御ユニット８００や、チャネル制御ユニット３００、ディスク制御ユニット４００、グローバルキャッシュ６００のいずれも装着することができるので、顧客ニーズに応じた柔軟性の高いストレージシステム１００構成することができる。これは、各ユニットのサイズやコネクタの位置、コネクタのピン配列等に互換性をもたせるようにしているからである。またボリウム管理テーブル８３５を設けることにより、異種ユニットが混在して装着されているディスク制御装置１１０においても、ホスト計算機２００から送信されるデータ入出力要求に対して、各ユニットにおいてデータ入出力制御を行うことができるからである。

【０１１９】

また本実施の形態に係るディスク制御装置１１０においては、キャッシュ領域管理テーブル８３３を設けることにより、ストレージシステム１００の性能向上を図ることができる。すなわちキャッシュ領域管理テーブル８３３の内容を変更することにより、ホスト計算機２００から受信するデータ入出力要求の特性に適合した、キャッシュ領域を確保することが可能となる。これにより例えば、ホスト計算機２００から受信したデータ入出力要求の多くが、当該データ入出力要求を受信したストレージ制御ユニット８００に接続された記憶ボリウムに対するものである場合には、自ＳＡＶＯＬ用領域８３６Ａの割り当てを増やすようにすることにより、ホスト計算機２００からのデータ入出力要求に対するキャッシュヒット率を増加させることができ、ストレージシステム１００の性能向上を図ることが可能となる。

【０１２０】

また本実施の形態に係るディスク制御装置１１０においては、グローバルキャッシュ６００にもボリウム管理テーブル６０５を設けることにより、例えばあるストレージ制御ユニット８００がホスト計算機２００からデータ入出力要求を受信した場合に、当該ストレージ制御ユニット８００のローカルキャッシュ８３０を参照しても、データ入出力要求の対象となっている記憶ボリウムに対するユニットが特定できなかった場合であっても、グローバルキャッシュ６００のボリウム管理テーブル６０５を参照することにより記憶ボリウムに対する入出力制御を行うユニットを特定することができるようになる。

【０１２１】

また他のストレージ制御ユニット８００に接続された記憶ボリウムに対するデータ入出力処理を行う場合には、ストレージ制御ユニット８００間で通信バッファ８３７を介してデータ入出力要求の授受を行うようにすることにより、各ストレージ制御ユニット８００は、他のストレージ制御ユニット８００の処理に依存せずに、当該他のストレージ制御ユニットに接続されている記憶ボリウムのデータに対するデータ入出力制御を行うことが可能となる。

【０１２２】

また、ストレージ制御ユニット８００に接続されている記憶ボリウムに記憶されているデータの複製を、ディスク制御ユニット４００に接続されている記憶ボリウムに書き込むことにより、それ以降のホスト計算機２００からのデータ入出力要求に対する入出力制御を当該ディスク制御ユニット４００に接続された記憶ボリウムに対して行うようにすることができる。このようにすることにより、ストレージシステム１００の構成変更の柔軟性を高めることができるようになる。例えば、ストレージシステム１００の導入時には初期コントローラ１１１により運用していた顧客が、その後のストレージシステム１００の規模拡大時に、チャネル制御ユニット３００とディスク制御ユニット４００とを用いた、拡張性の高いシステム構成に変更するようにすることが出来るようになる。

【０１２３】

以上本実施の形態について説明したが、上記実施例は本発明の理解を容易にす

るためのものであり、本発明を限定して解釈するためのものではない。本発明はその趣旨を逸脱することなく変更、改良され得ると共に、本発明にはその等価物も含まれる。

【０１２４】

【発明の効果】

記憶デバイス制御装置、及び記憶デバイス制御装置の制御方法を提供することができる。

【図面の簡単な説明】

【図１】　本実施の形態に係るストレージシステムの外観構成を示す図である。

【図２】　本実施の形態に係るストレージシステムの全体構成を示すブロック図である。

【図３】　本実施の形態に係るストレージ制御ユニットを示す図である。

【図４】　本実施の形態に係るチャネル制御ユニットを示す図である。

【図５】　本実施の形態に係るディスク制御ユニットを示す図である。

【図６】　本実施の形態に係るキャッシュメモリユニットを示す図である。

【図７】　本実施の形態に係るストレージシステムに、ストレージ制御ユニット、チャネル制御ユニット、ディスク制御ユニット、又はキャッシュメモリユニットが装着される様子を示す図である。

【図８】　本実施の形態に係るストレージ制御ユニットの機能を示すブロック図である。

【図９】　本実施の形態に係るストレージ制御ユニット間を接続するペア間接続部を説明するためのブロック図である。

【図１０】　本実施の形態に係るチャネル制御ユニットの機能を示すブロック図である。

【図１１】　本実施の形態に係るチャネル制御ユニット間を接続するペア間接続部を説明するためのブロック図である。

【図１２】　本実施の形態に係るディスク制御ユニットの機能を示すブロック図である。

【図１３】　本実施の形態に係るローカルキャッシュメモリの構成を示すブロック図である。

【図１４】　本実施の形態に係るグローバルキャッシュメモリの構成を示すブロック図である。

【図１５】　本実施の形態に係る内部接続部の構成を示すブロック図である。

【図１６】　本実施の形態に係る管理端末の構成を示すブロック図である。

【図１７】　本実施の形態に係るストレージシステムの規模を拡大する際の外観構成の変化の一例を示す図である。

【図１８】　本実施の形態に係るストレージシステムの規模を拡大する前の構成の一例を示す図である。

【図１９】　本実施の形態に係るストレージシステムの規模を拡大する前の構成の一例を示す図である。

【図２０】　本実施の形態に係るストレージシステムの規模を拡大した後の構成の一例を示す図である。

【図２１】　本実施の形態に係るストレージシステムの規模を拡大する際の外観構成の変化の一例を示す図である。

【図２２】　本実施の形態に係るストレージシステムにおける初期コントローラを示す図である。

【図２３】　本実施の形態に係るストレージシステムの規模を拡大する際の構成の一例を示す図である。

【図２４】　本実施の形態に係るストレージシステムの規模を拡大する際の外観構成の変化の一例を示す図である。

【図２５】　本実施の形態に係るストレージシステムの規模を拡大する際の構成の一例を示す図である。

【図２６】　本実施の形態に係るストレージシステムにおいて、ボリウム管理テーブルをローカルキャッシュメモリからグローバルキャッシュメモリへ移行する処理を示すフローチャートである。

【図２７】　本実施の形態に係るストレージシステムにおいて、ボリウムを

作成した場合に行われるボリウム管理テーブルを更新する処理を示すフローチャートである。

【図２８】　本実施の形態に係るデータアクセス処理を示すフローチャートである。

【図２９】　本実施の形態に係るローカルキャッシュメモリのヒットミス判定処理を示すフローチャートである。

【図３０】　本実施の形態に係るグローバルキャッシュメモリのロック確保処理を示すフローチャートである。

【図３１】　本実施の形態に係るローカルキャッシュメモリのステージング処理を示すフローチャートである。

【図３２】　本実施の形態に係るローカルキャッシュメモリのデステージング処理を示すフローチャートである。

【図３３】　本実施の形態に係るリードライト処理を示すフローチャートである。

【図３４】　本実施の形態に係るローカルキャッシュメモリへのライト処理を示すフローチャートである。

【図３５】　本実施の形態に係るメッセージ通信を送信する側によるデータアクセス処理を示すフローチャートである。

【図３６】　本実施の形態に係るメッセージ通信を受信する側によるデータアクセス処理を示すフローチャートである。

【図３７】　本実施の形態に係るコマンドの構成を示す図である。

【図３８】　本実施の形態に係るコマンドを示す図である。

【図３９】　本実施の形態に係るメッセージを示す図である。

【図４０】　本実施の形態に係るコマンドの送受信を示す図である。

【図４１】　本実施の形態に係るメッセージの送受信を示す図である。

【図４２】　本実施の形態に係るアクセス方法を変更する場合の処理を示す図である。

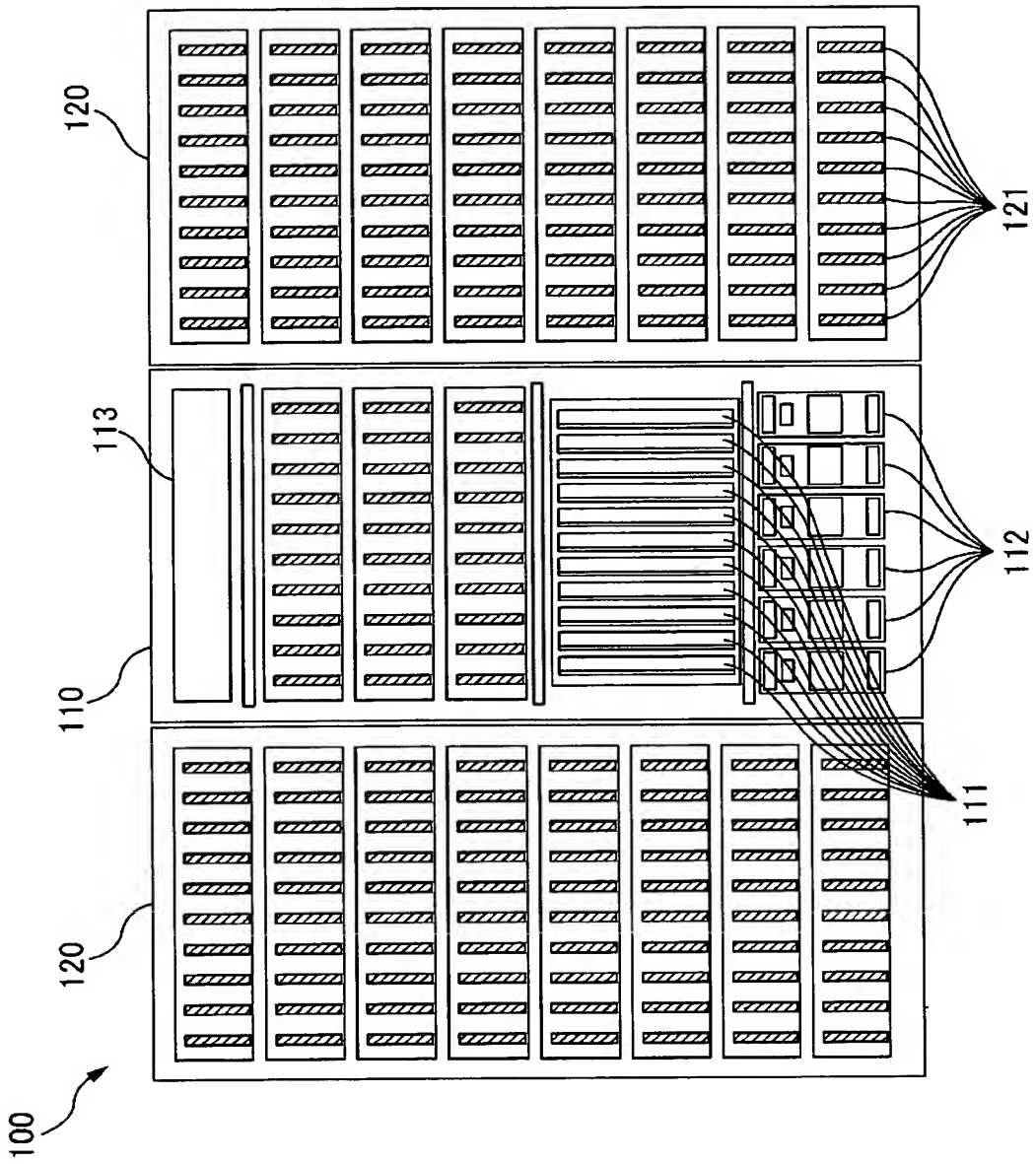
【符号の説明】

１００　ストレージシステム　　１１０　ディスク制御装置

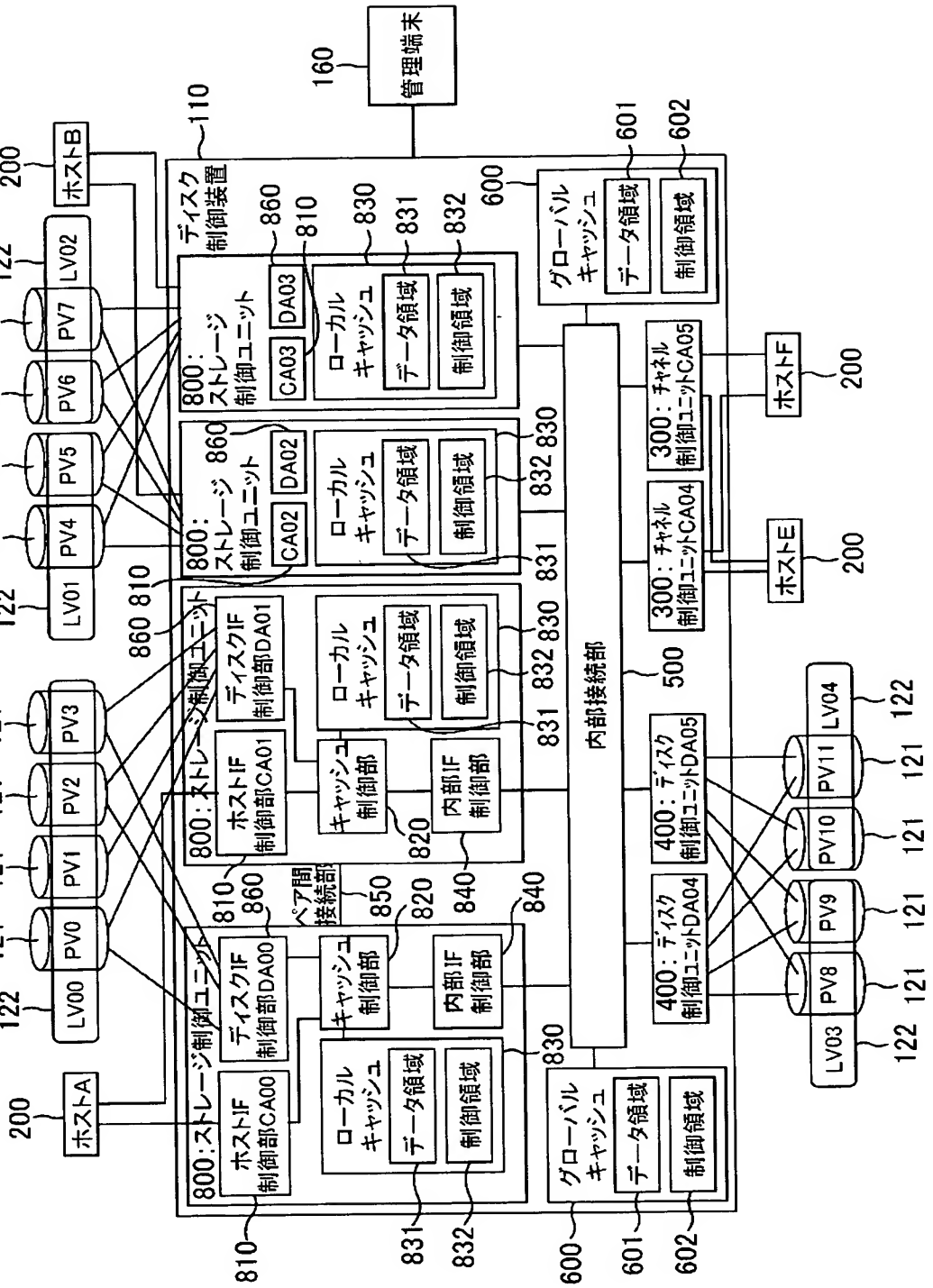
１２０　ディスク駆動装置　　１３０　装着部
１６０　管理端末　　２００　ホスト計算機
３００　チャネル制御ユニット　　３１０　ホストＩＦ制御部
３３０　ローカルキャッシュ　　３３１　データ領域
３３２　制御領域　　４００　ディスク制御ユニット
４６０　ディスクＩＦ制御部　　５００　内部接続部
６００　グローバルキャッシュ　　６０３　キャッシュ領域管理テーブル
６０４　キャッシュデータ管理テーブル　６０５　ボリウム管理テーブル
６０６　ダイレクトアクセス用データ領域　６０７　通信バッファ
８００　ストレージ制御ユニット　　８１０　ホストＩＦ制御部
８３０　ローカルキャッシュ　　８３３　キャッシュ領域管理テーブル
８３４　キャッシュデータ管理テーブル　８３５　ボリウム管理テーブル
８３６　ダイレクトアクセス用データ領域
８３７　通信バッファ　　８６０　ディスクＩＦ制御部

【書類名】　図面

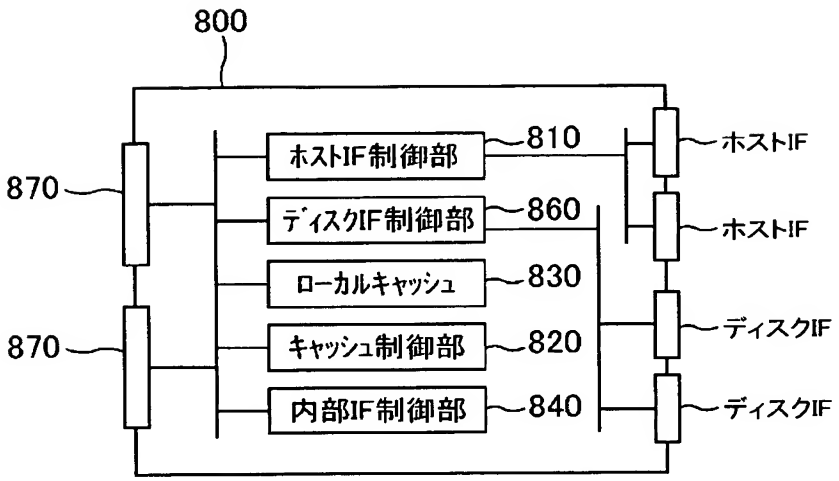
【図１】



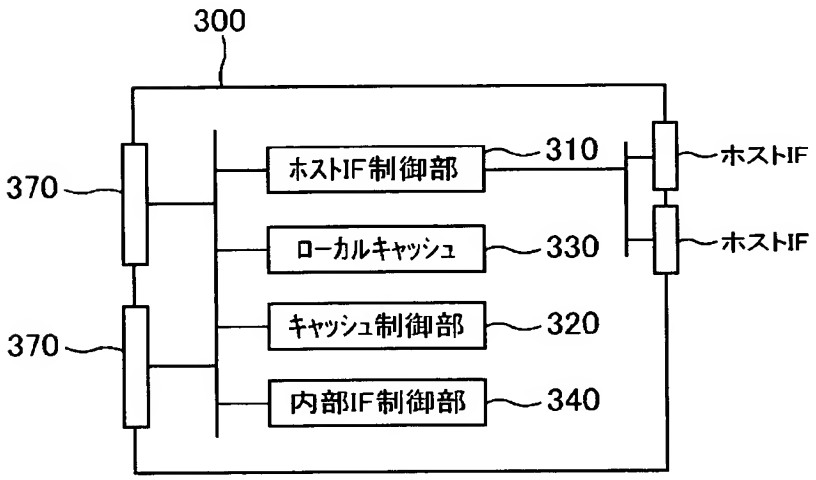
【図２】


管理端末
160
ディスク制御装置
110
ホストA
ホストB
ホストE
ホストF
200
LV00
LV01
LV02
LV03
LV04
PV0
PV1
PV2
PV3
PV4
PV5
PV6
PV7
PV8
PV9
PV10
PV11
121
122
800:ストレージ制御ユニット
ホストIF制御部CA00
ディスクIF制御部DA00
ホストIF制御部CA01
ディスクIF制御部DA01
800:ストレージ制御ユニット
CA02
DA02
CA03
DA03
ローカルキャッシュ
データ領域
制御領域
キャッシュ制御部
内部IF制御部
ペア間接続部
グローバルキャッシュ
300:チャネル制御ユニットCA04
300:チャネル制御ユニットCA05
400:ディスク制御ユニットDA04
400:ディスク制御ユニットDA05
内部接続部
500
600
601
602
810
820
830
831
832
840
850
860

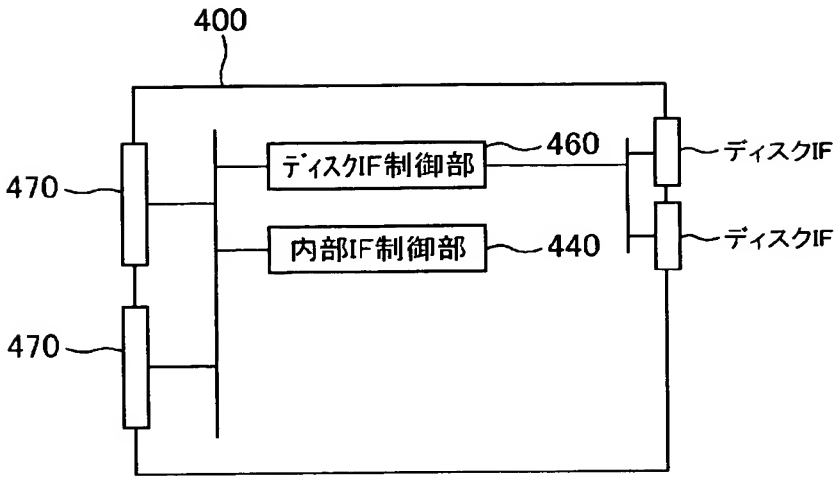
【図３】


800
ホストIF制御部
810
ホストIF
870
ディスクIF制御部
860
ホストIF
ローカルキャッシュ
830
ディスクIF
870
キャッシュ制御部
820
内部IF制御部
840
ディスクIF

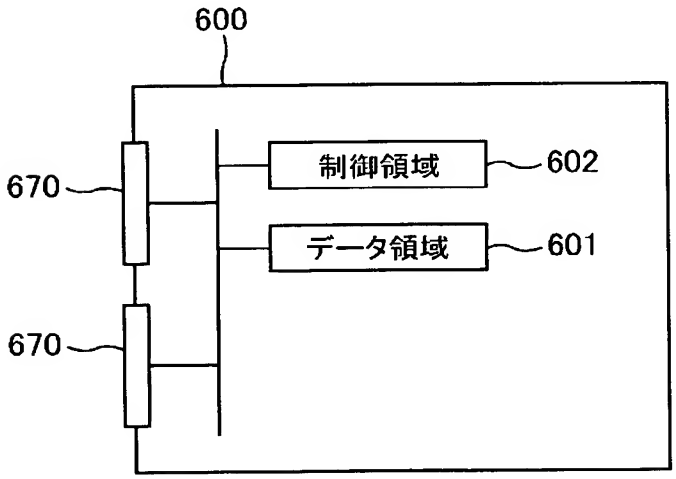
【図４】


300
ホストIF制御部
310
ホストIF
370
ローカルキャッシュ
330
ホストIF
キャッシュ制御部
320
370
内部IF制御部
340

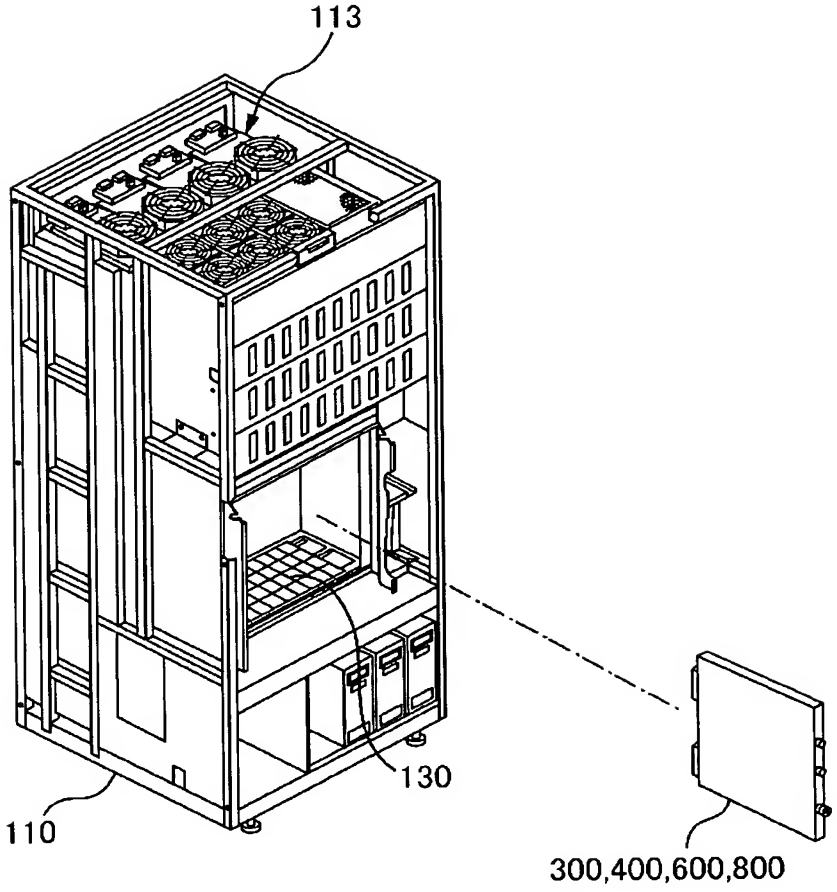
【図５】


400
470
ディスクIF制御部
460
ディスクIF
内部IF制御部
440
ディスクIF
470

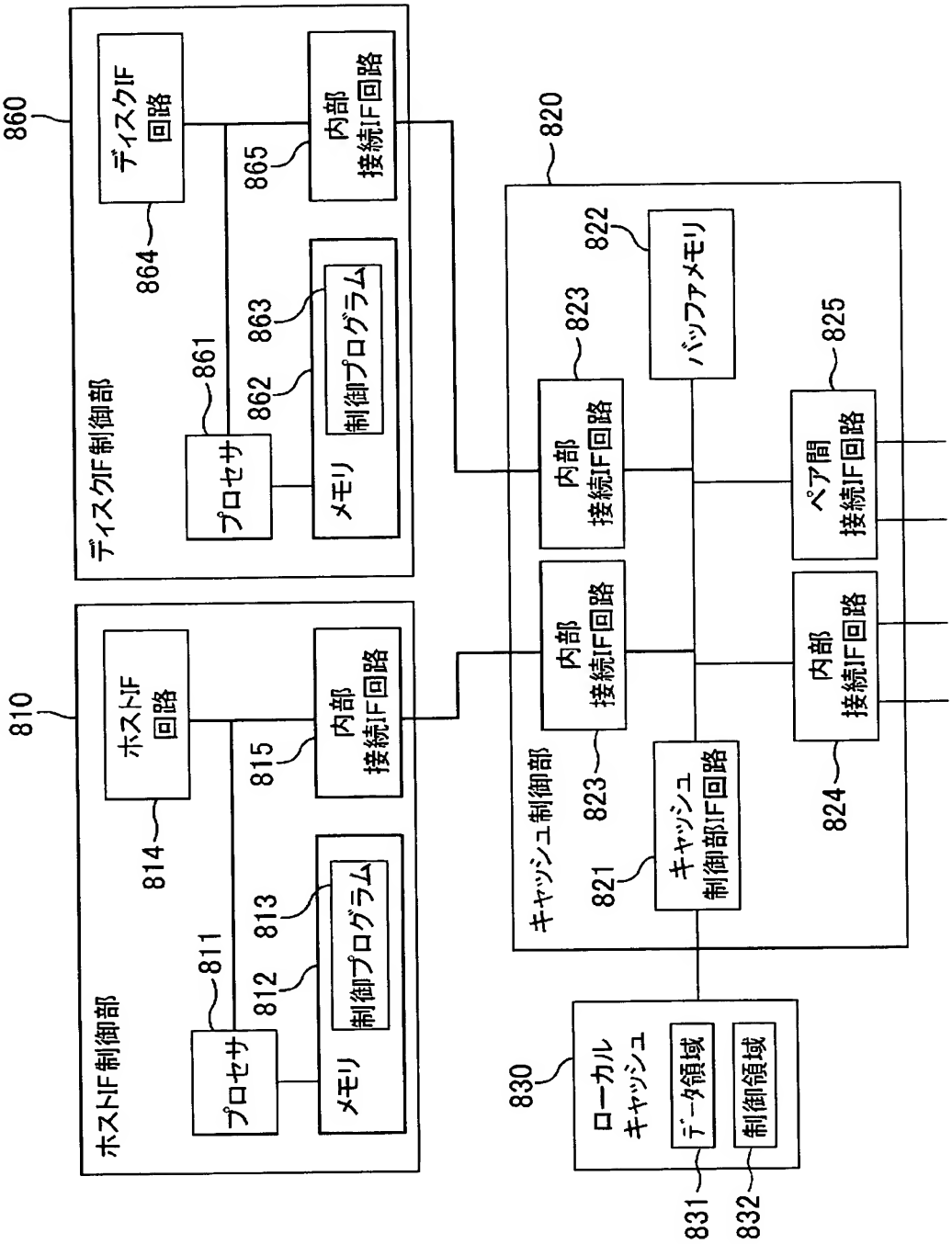
【図６】


600
670
制御領域
602
データ領域
601
670

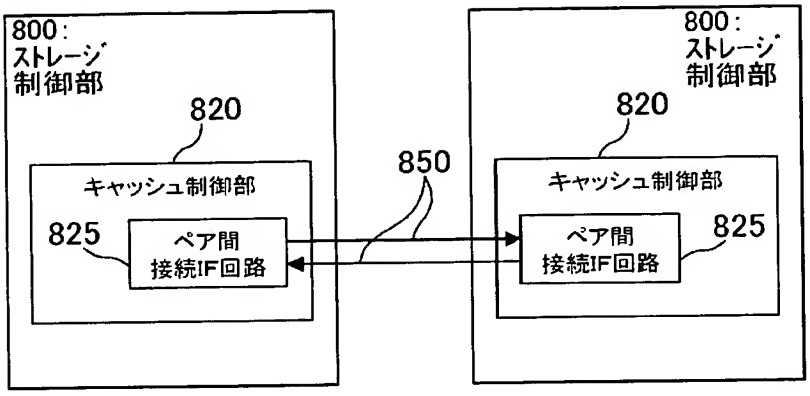
【図７】



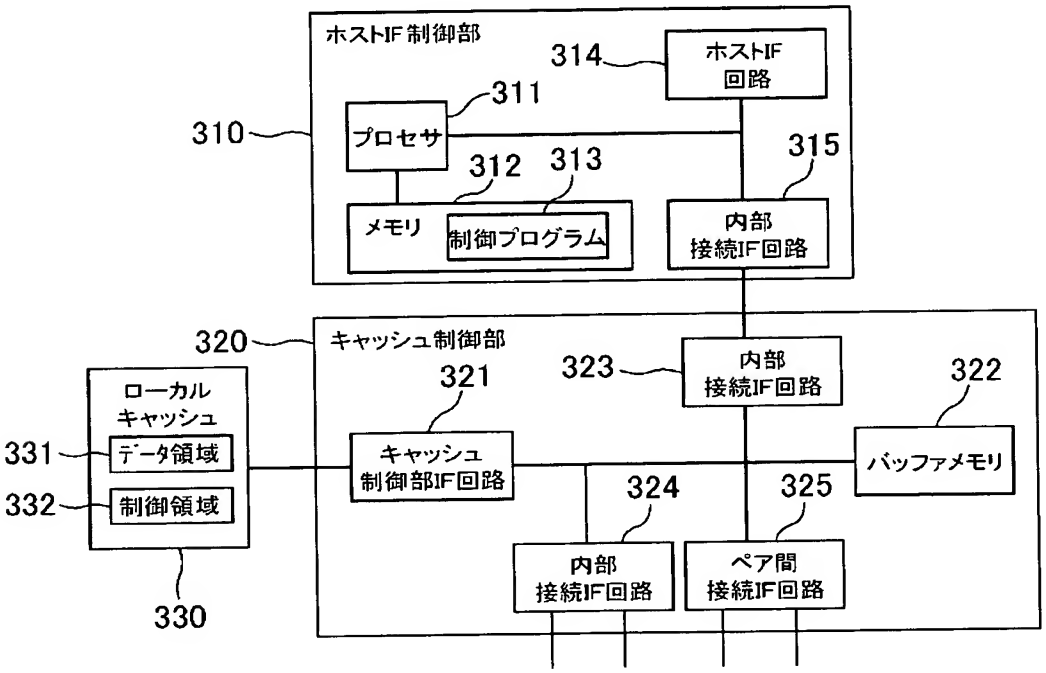
【図８】


860
ディスクIF制御部
ディスクIF回路
内部接続IF回路
プロセサ
メモリ
制御プログラム
861
862
863
864
865
810
ホストIF制御部
ホストIF回路
内部接続IF回路
プロセサ
メモリ
制御プログラム
811
812
813
814
815
820
キャッシュ制御部
内部接続IF回路
内部接続IF回路
キャッシュ制御部IF回路
バッファメモリ
ペア間接続IF回路
内部接続IF回路
821
822
823
823
824
825
830
ローカルキャッシュ
データ領域
制御領域
831
832

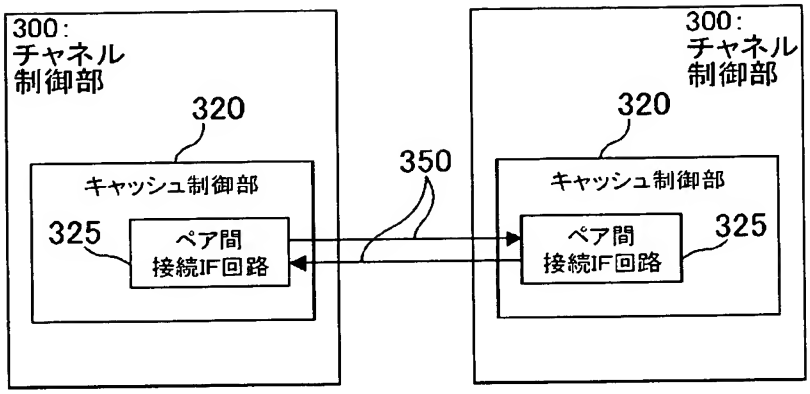
【図９】


800:
ストレージ
制御部
820
キャッシュ制御部
825
ペア間
接続IF回路
850
800:
ストレージ
制御部
820
キャッシュ制御部
ペア間
接続IF回路
825

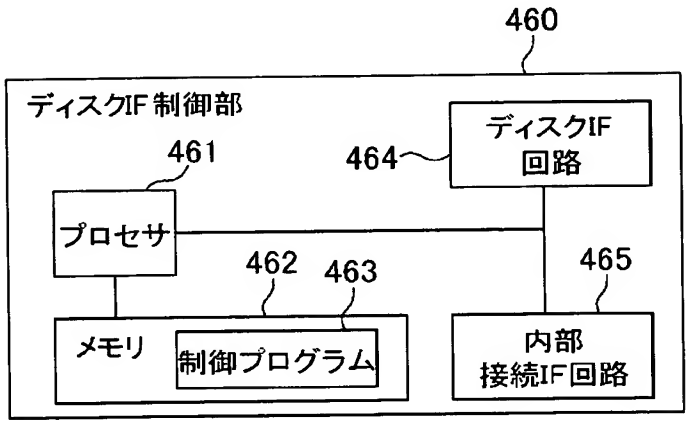
【図１０】


ホストIF制御部
314
ホストIF
回路
311
310
プロセサ
315
312
313
メモリ
制御プログラム
内部
接続IF回路
320
キャッシュ制御部
323
内部
接続IF回路
321
322
ローカル
キャッシュ
331
データ領域
キャッシュ
制御部IF回路
バッファメモリ
332
制御領域
324
325
内部
接続IF回路
ペア間
接続IF回路
330

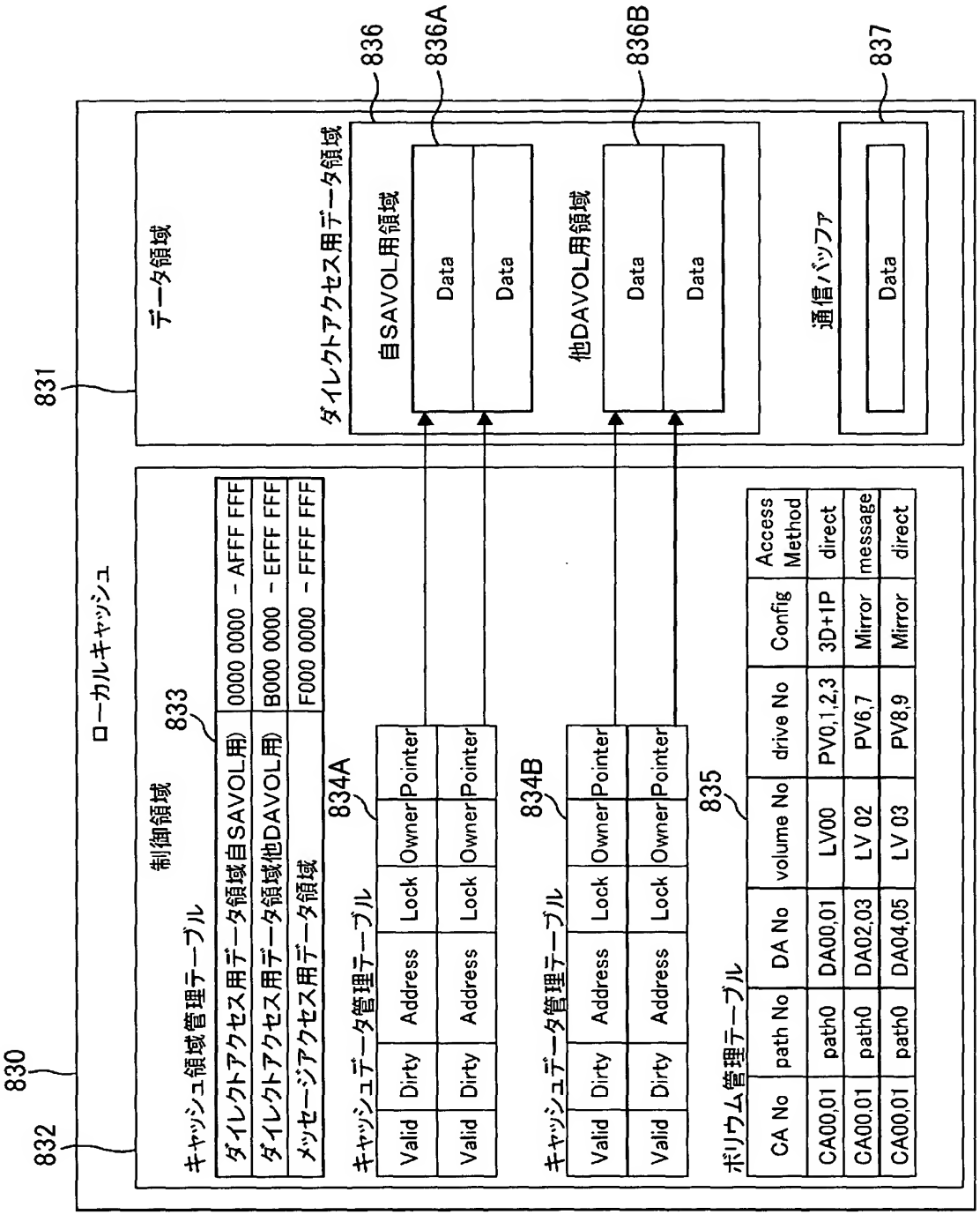
【図１１】


300:
チャネル
制御部
320
キャッシュ制御部
325
ペア間
接続IF回路
350
300:
チャネル
制御部
320
キャッシュ制御部
ペア間
接続IF回路
325

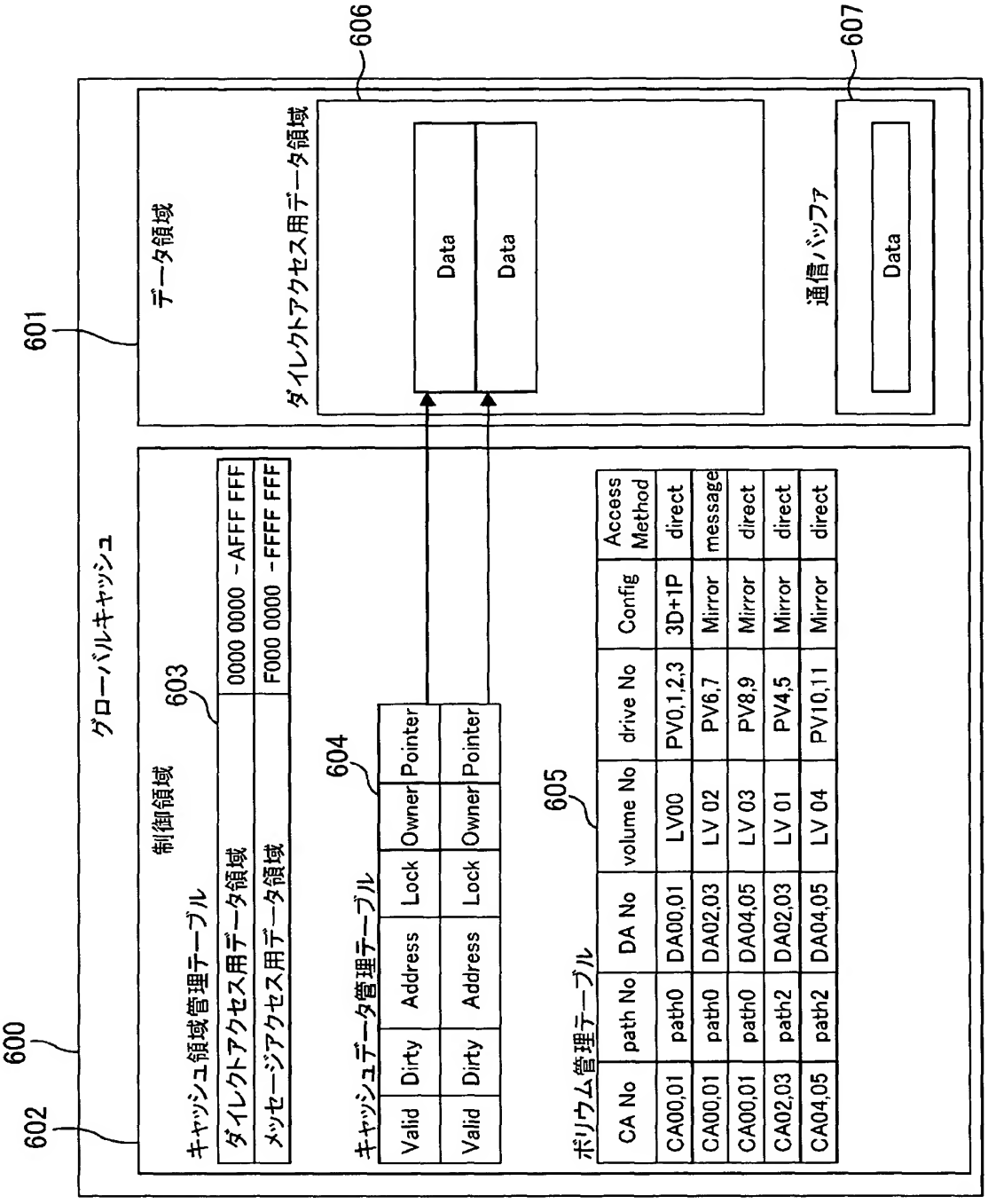
【図１２】


460
ディスクIF制御部
461
464
ディスクIF
回路
プロセサ
462
463
465
メモリ
制御プログラム
内部
接続IF回路

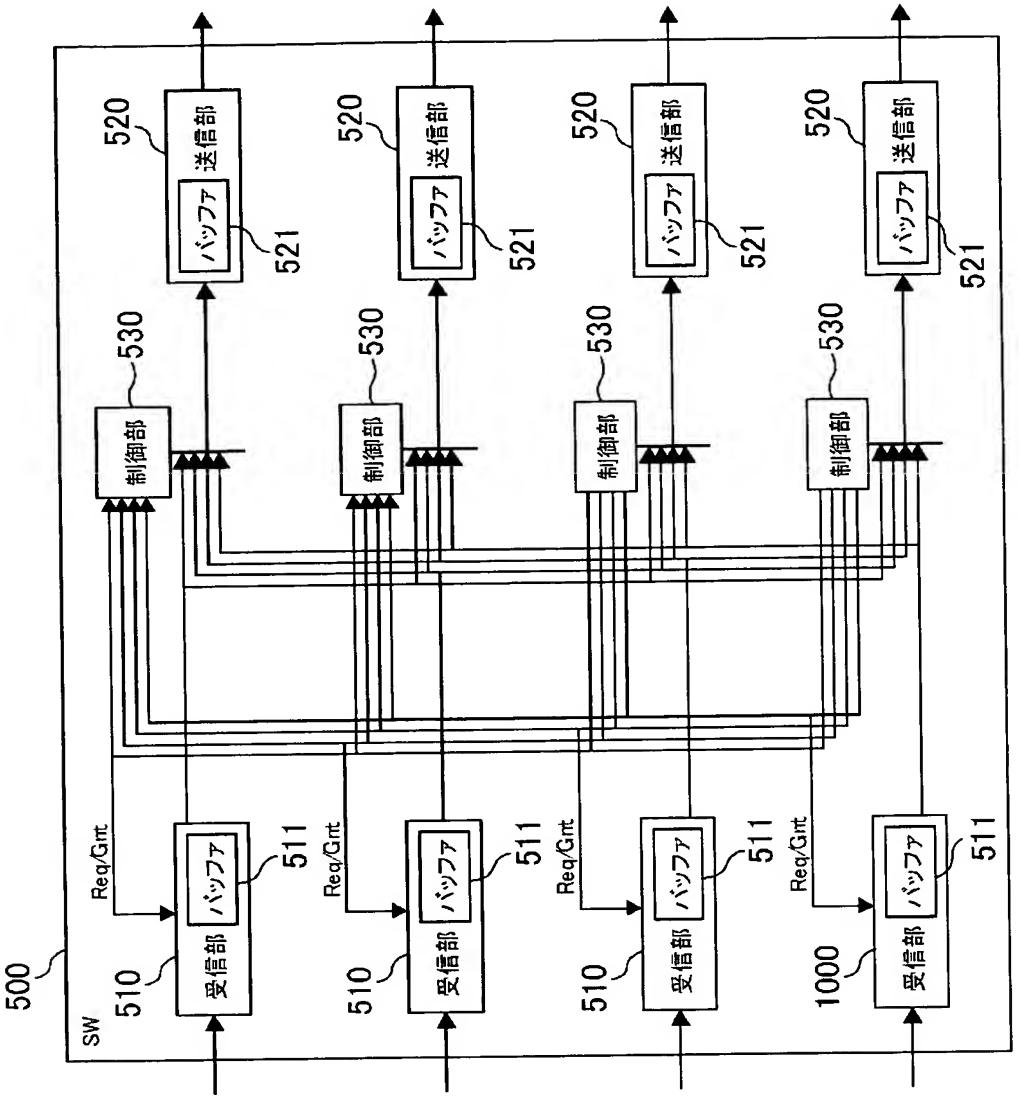
【図１３】


830
832
ローカルキャッシュ
831
データ領域
ダイレクトアクセス用データ領域
836
836A
自SAVOL用領域
Data
Data
836B
他DAVOL用領域
Data
Data
837
通信バッファ
Data
制御領域
833
キャッシュ領域管理テーブル
ダイレクトアクセス用データ領域自SAVOL用)
ダイレクトアクセス用データ領域他DAVOL用)
メッセージアクセス用データ領域
0000 0000 - AFFF FFF
B000 0000 - EFFF FFF
F000 0000 - FFFF FFF
834A
キャッシュデータ管理テーブル
Valid
Dirty
Address
Lock
Owner
Pointer
834B
キャッシュデータ管理テーブル
835
ボリューム管理テーブル
CA No
path No
DA No
volume No
drive No
Config
Access Method
CA00,01
path0
DA00,01
LV00
PV0,1,2,3
3D+1P
direct
CA00,01
path0
DA02,03
LV 02
PV6,7
Mirror
message
CA00,01
path0
DA04,05
LV 03
PV8,9
Mirror
direct

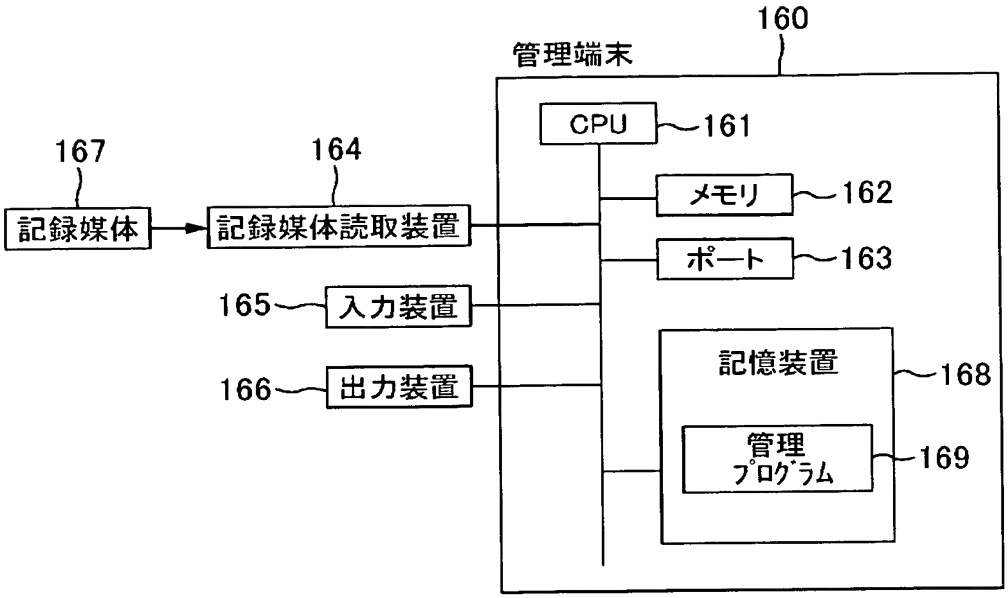
【図１４】


600
601
602
グローバルキャッシュ
制御領域
データ領域
603
キャッシュ領域管理テーブル
ダイレクトアクセス用データ領域 0000 0000 -AFFF FFF
メッセージアクセス用データ領域 F000 0000 -FFFF FFF
604
キャッシュデータ管理テーブル
Valid Dirty Address Lock Owner Pointer
Valid Dirty Address Lock Owner Pointer
605
ボリウム管理テーブル
CA No | path No | DA No | volume No | drive No | Config | Access Method
CA00,01 | path0 | DA00,01 | LV00 | PV0,1,2,3 | 3D+1P | direct
CA00,01 | path0 | DA02,03 | LV 02 | PV6,7 | Mirror | message
CA00,01 | path0 | DA04,05 | LV 03 | PV8,9 | Mirror | direct
CA02,03 | path2 | DA02,03 | LV 01 | PV4,5 | Mirror | direct
CA04,05 | path2 | DA04,05 | LV 04 | PV10,11 | Mirror | direct
606
ダイレクトアクセス用データ領域
Data
Data
607
通信バッファ
Data

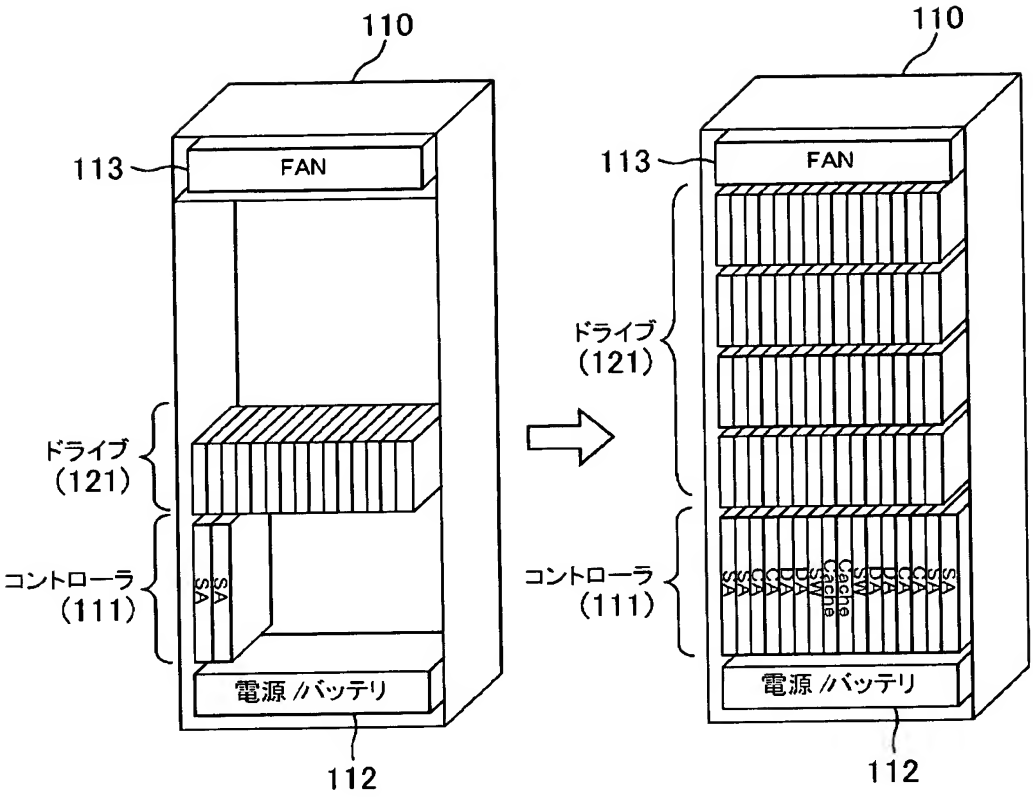
【図１５】



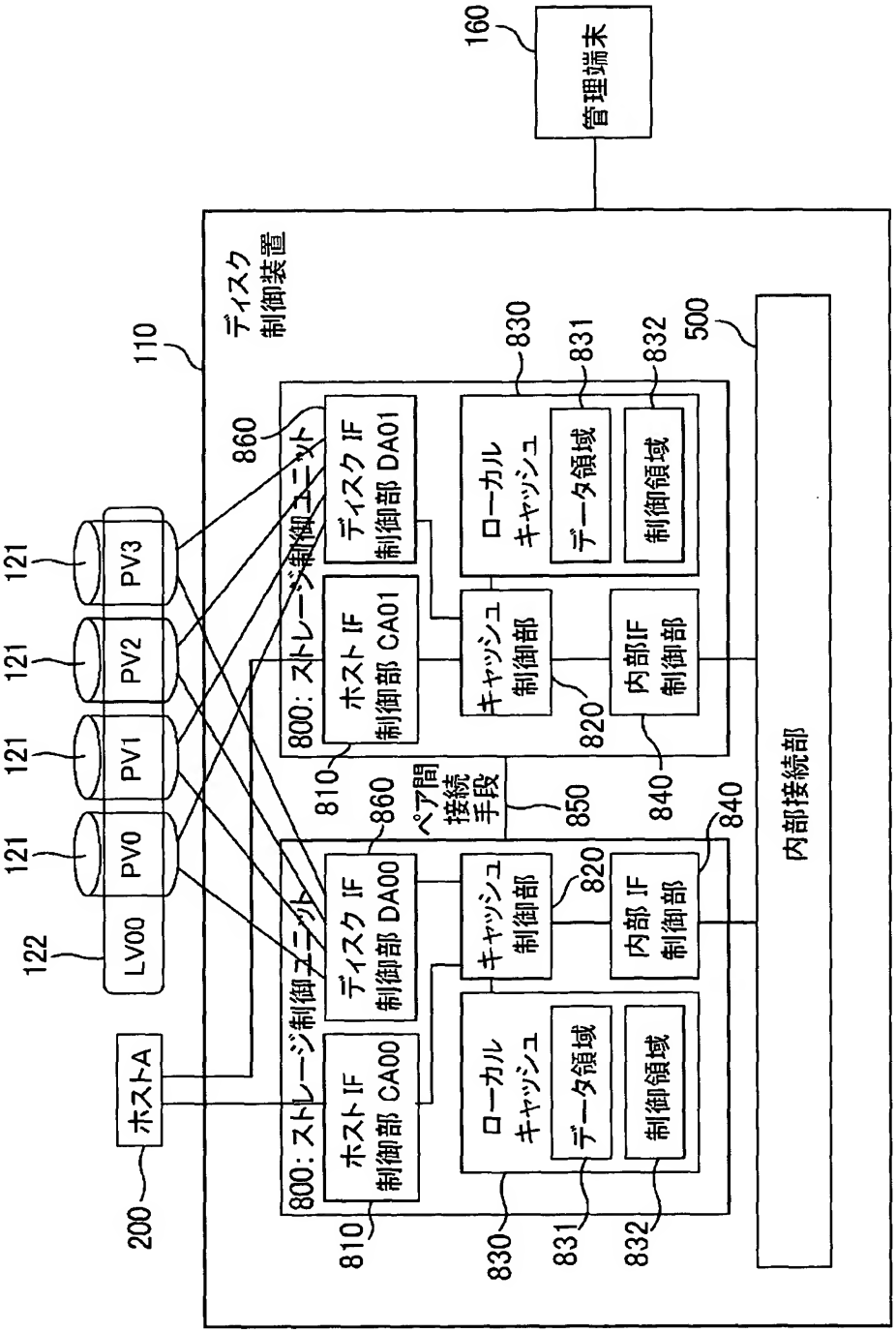
【図１６】


160
管理端末
CPU
161
167
164
記録媒体
記録媒体読取装置
メモリ
162
ポート
163
165
入力装置
記憶装置
168
166
出力装置
管理
プログラム
169

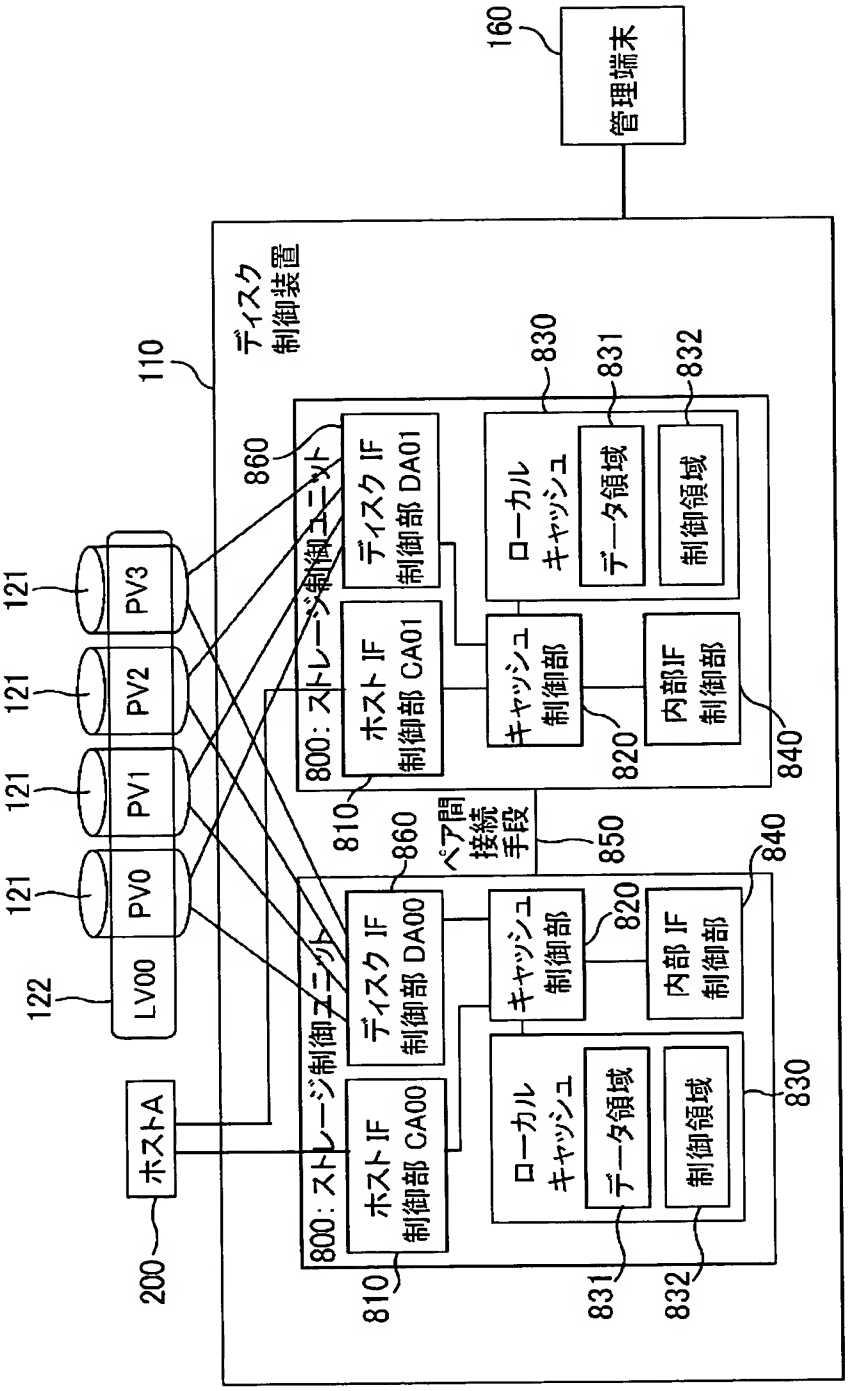
【図１７】


110
113
FAN
ドライブ
(121)
コントローラ
(111)
SA
SA
電源 /バッテリ
112
110
113
FAN
ドライブ
(121)
コントローラ
(111)
Cache
電源 /バッテリ
112

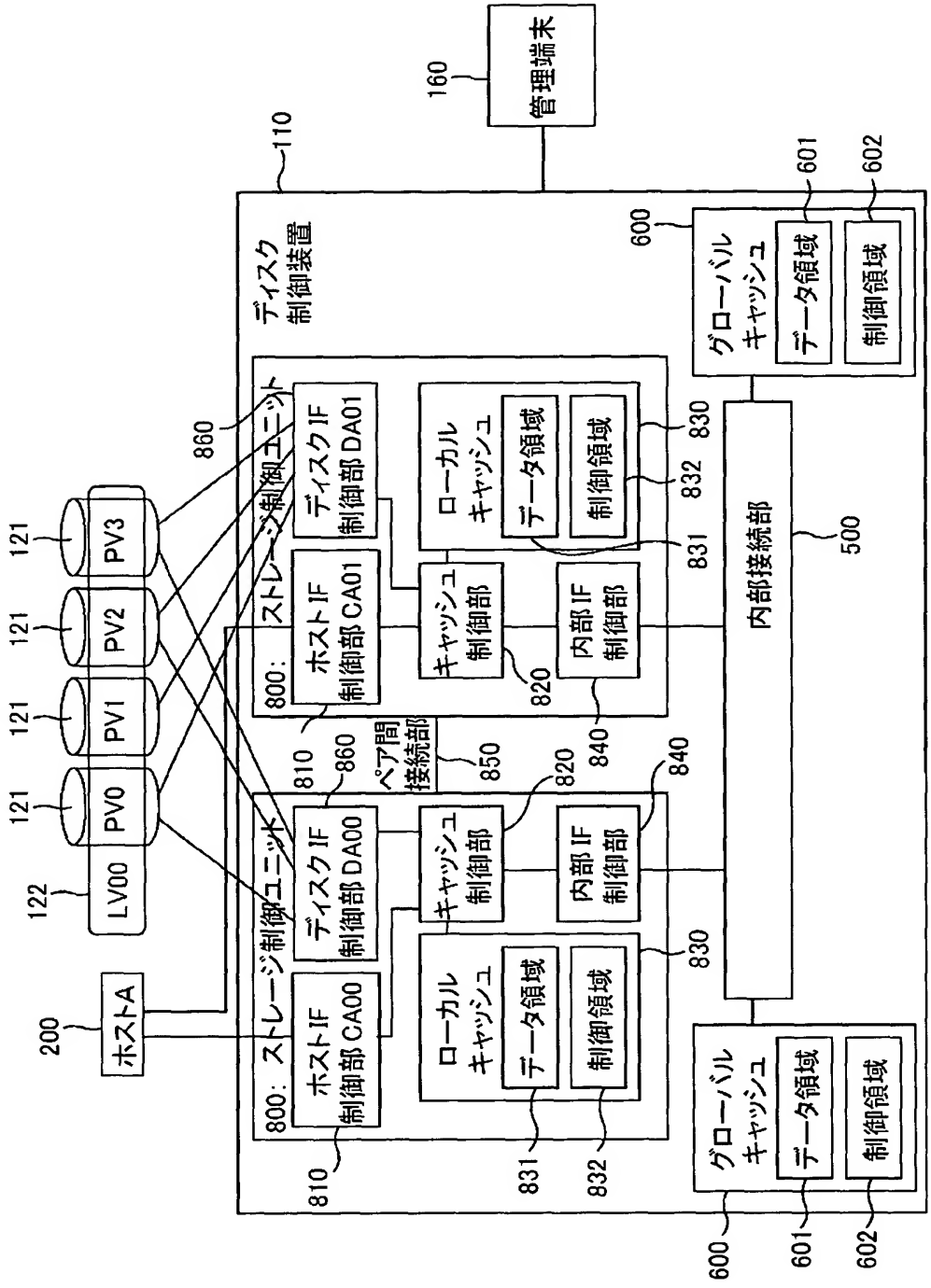
【図１８】


160
管理端末
ディスク
制御装置
110
500
内部接続部
830
831
832
ローカル
キャッシュ
データ領域
制御領域
860
ディスク IF
制御部 DA01
800: ストレージ制御ユニット
ホスト IF
制御部 CA01
キャッシュ
制御部
820
内部IF
制御部
840
810
ペア間
接続
手段
850
121
PV3
PV2
PV1
PV0
122
LV00
ディスク IF
制御部 DA00
ホスト IF
制御部 CA00
ホストA
200

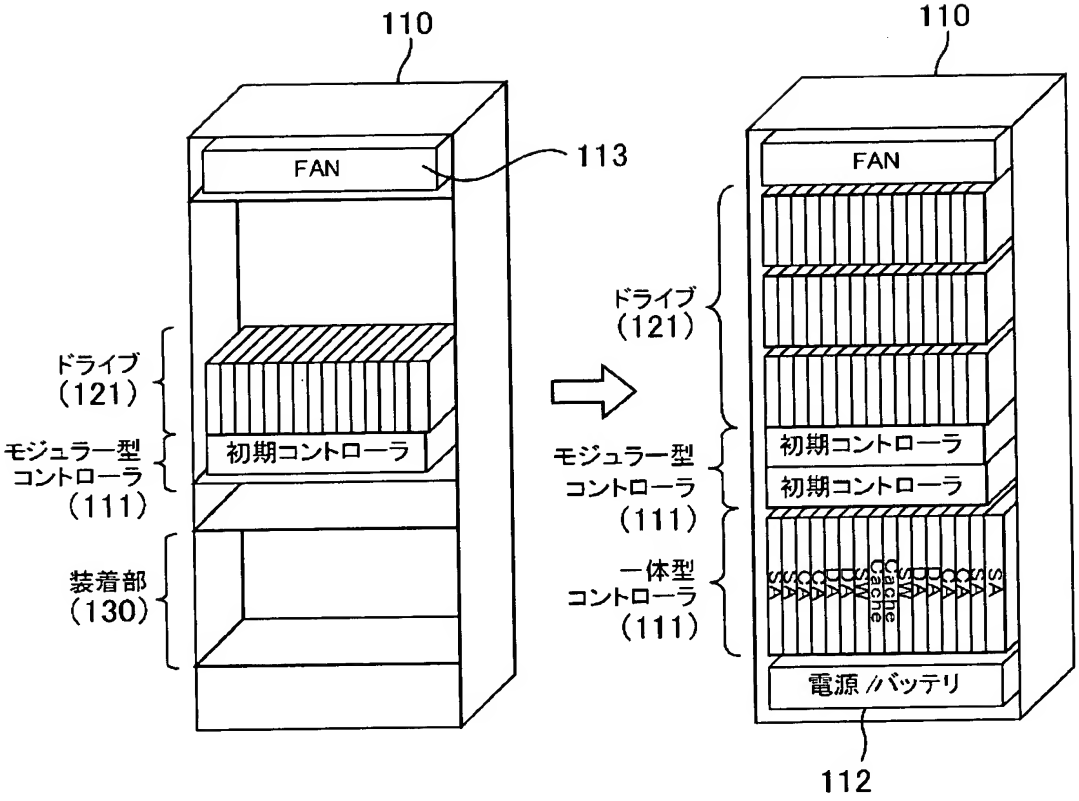
【図１９】


160
管理端末
ディスク
制御装置
110
830
831
832
860
ディスク IF
制御部 DA01
ローカル
キャッシュ
データ領域
制御領域
800: ストレージ制御ユニット
ホスト IF
制御部 CA01
キャッシュ
制御部
内部IF
制御部
820
840
121
PV3
121
PV2
121
PV1
121
PV0
122
LV00
810
ペア間
接続
手段
850
860
840
ディスク IF
制御部 DA00
キャッシュ
制御部
820
内部 IF
制御部
ホストA
800: ストレージ制御ユニット
ホスト IF
制御部 CA00
ローカル
キャッシュ
データ領域
制御領域
830
200
810
831
832

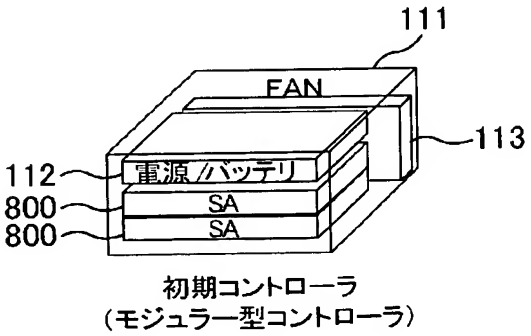
【図２０】


管理端末
160
ディスク制御装置
110
ストレージ制御ユニット
800:
ホストIF制御部 CA00
ホストIF制御部 CA01
ディスクIF制御部 DA00
ディスクIF制御部 DA01
キャッシュ制御部
内部IF制御部
ペア間接続部
850
ローカルキャッシュ
データ領域
制御領域
830
831
832
840
860
810
820
内部接続部
500
グローバルキャッシュ
600
601
602
PV0
PV1
PV2
PV3
LV00
121
122
ホストA
200

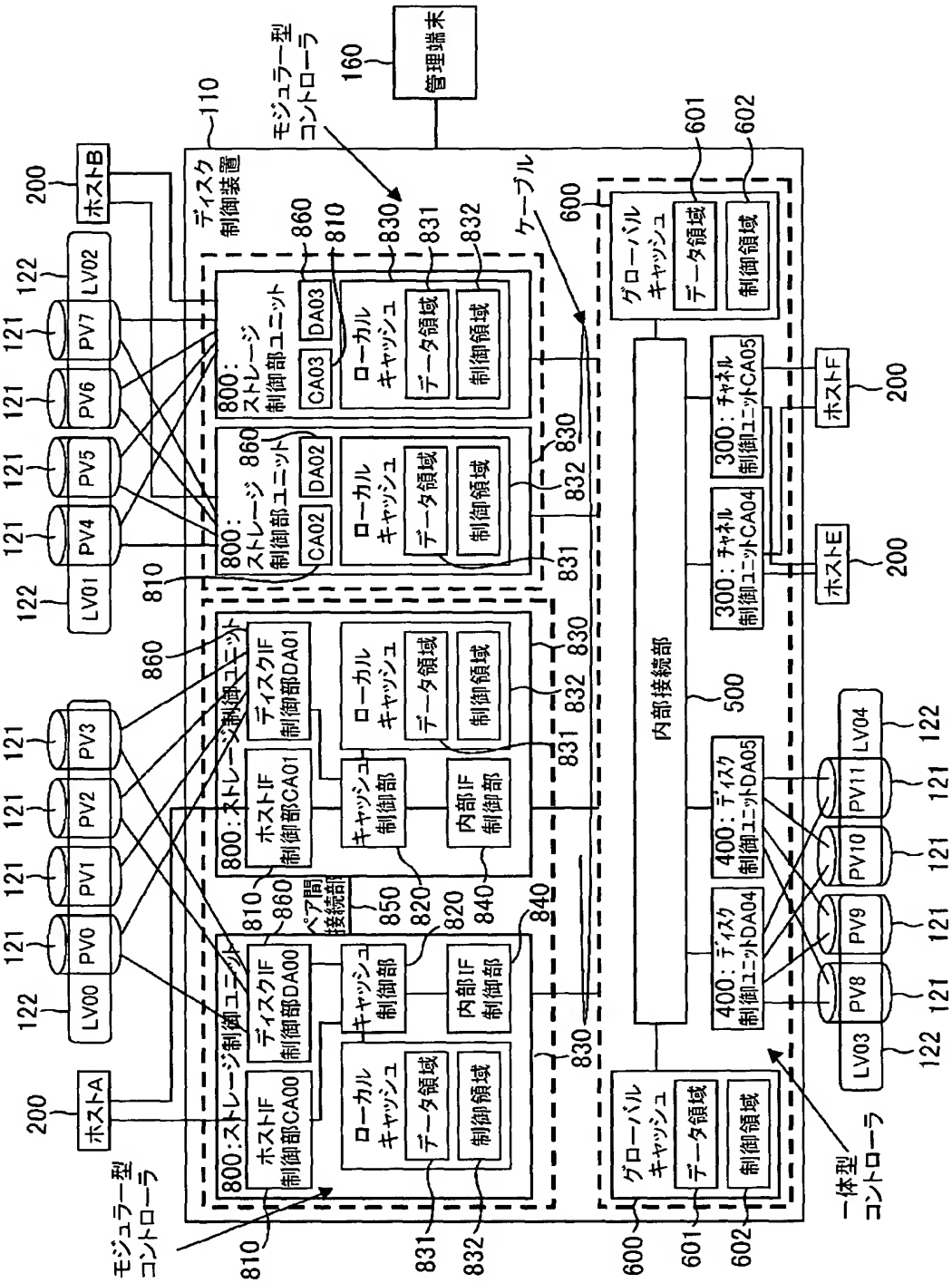
【図２１】


110
FAN
113
ドライブ
(121)
モジュラー型
コントローラ
(111)
初期コントローラ
装着部
(130)
110
FAN
ドライブ
(121)
モジュラー型
コントローラ
(111)
初期コントローラ
初期コントローラ
一体型
コントローラ
(111)
電源 /バッテリ
112

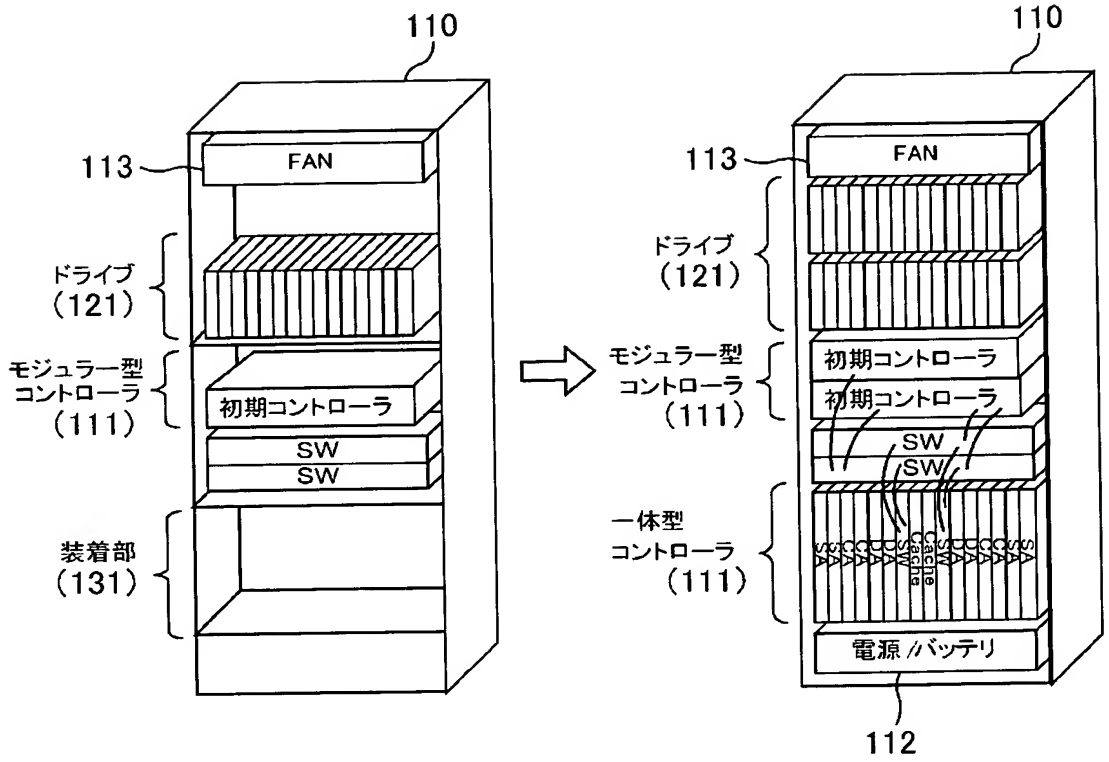
【図２２】


111
FAN
113
112
電源 /バッテリ
800
SA
800
SA
初期コントローラ
(モジュラー型コントローラ)

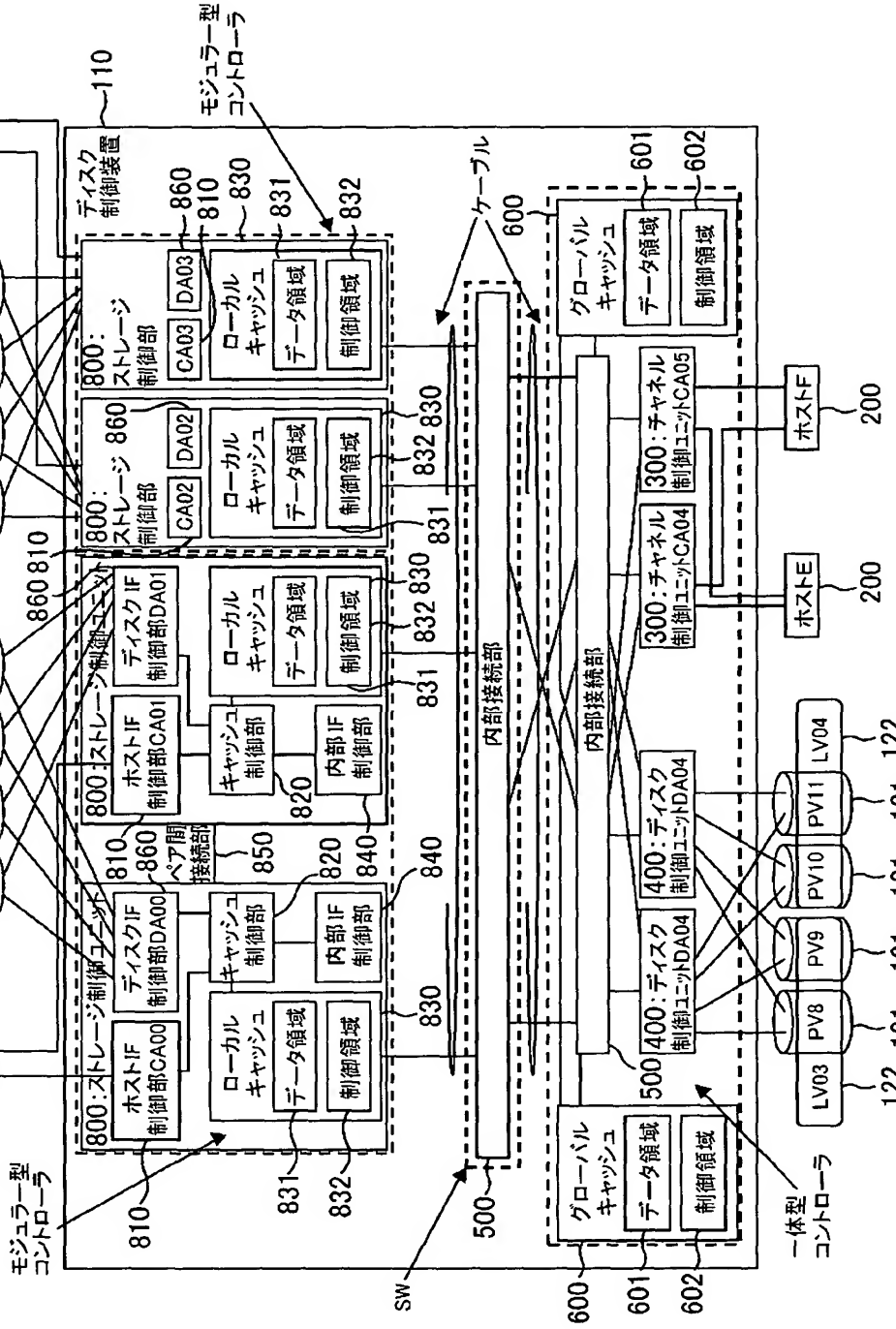
【図２３】


ディスク制御装置
110
モジュラー型コントローラ
160
管理端末
ケーブル
600
ホストA
ホストB
ホストE
ホストF
200
LV00
LV01
LV02
LV03
LV04
PV0
PV1
PV2
PV3
PV4
PV5
PV6
PV7
PV8
PV9
PV10
PV11
121
122
800:ストレージ制御ユニット
ホストIF制御部CA00
ディスクIF制御部DA00
ホストIF制御部CA01
ディスクIF制御部DA01
キャッシュ制御部
内部IF制御部
ペア間接続部
850
820
840
810
860
830
831
832
800:ストレージ制御部ユニット
CA02
CA03
DA02
DA03
ローカルキャッシュ
データ領域
制御領域
グローバルキャッシュ
601
602
内部接続部
500
300:チャネル制御ユニットCA04
300:チャネル制御ユニットCA05
400:ディスク制御ユニットDA04
400:ディスク制御ユニットDA05
一体型コントローラ

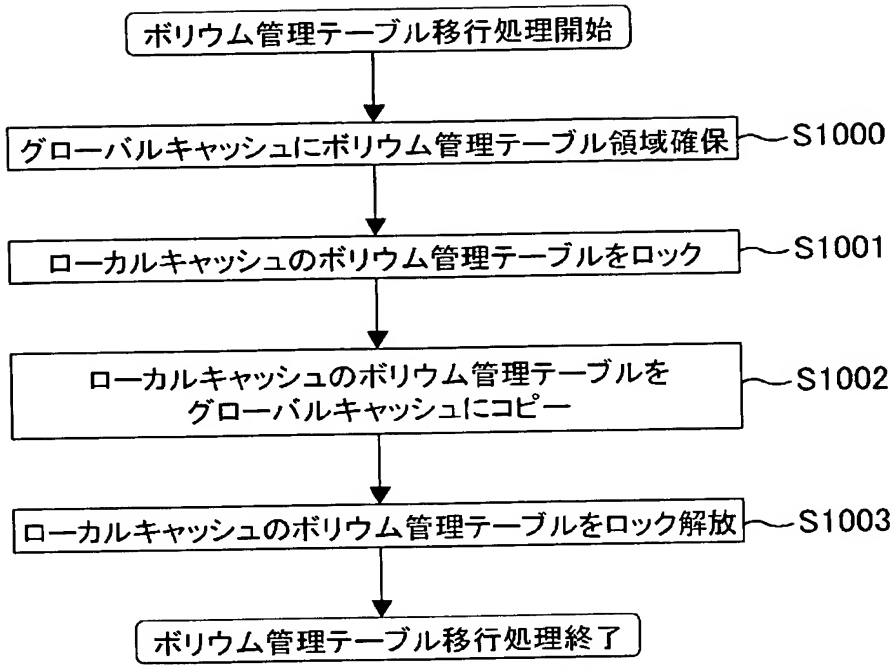
【図２４】


110
113
FAN
ドライブ
(121)
モジュラー型
コントローラ
(111)
初期コントローラ
SW
SW
装着部
(131)
110
113
FAN
ドライブ
(121)
モジュラー型
コントローラ
(111)
初期コントローラ
初期コントローラ
SW
SW
一体型
コントローラ
(111)
Cache
Cache
電源/バッテリ
112

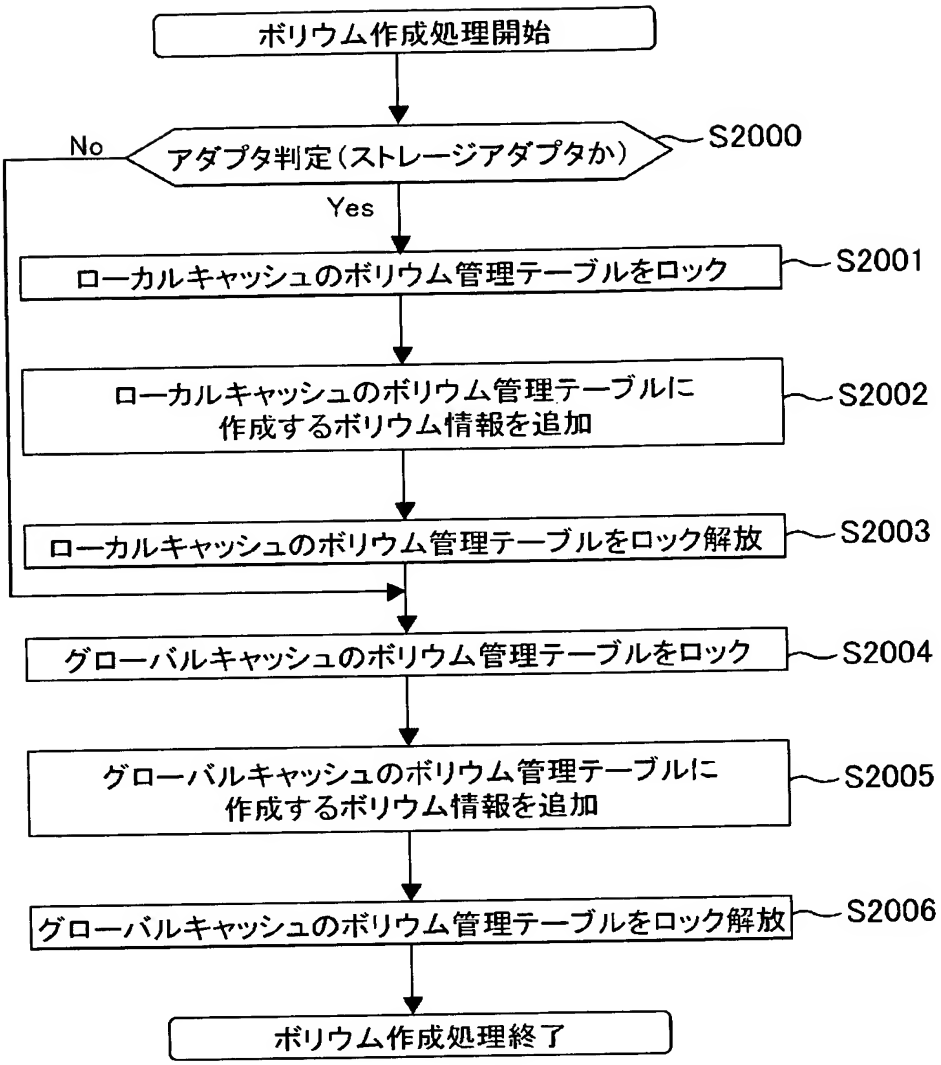
【図２５】


ホストA
ホストB
ホストE
ホストF
200
LV00
LV01
LV02
LV03
LV04
122
PV0
PV1
PV2
PV3
PV4
PV5
PV6
PV7
PV8
PV9
PV10
PV11
121
ディスク制御装置
110
モジュラー型コントローラ
800：ストレージ制御部
ホストIF制御部CA00
ディスクIF制御部DA00
ディスクIF制御部DA01
ホストIF制御部CA01
キャッシュ制御部
内部IF制御部
ペア間接続部
850
ローカルキャッシュ
データ領域
制御領域
810
820
830
840
860
CA02
CA03
DA02
DA03
831
832
ケーブル
600
内部接続部
500
SW
グローバルキャッシュ
601
602
300：チャネル制御ユニットCA04
300：チャネル制御ユニットCA05
400：ディスク制御ユニットDA04
400：ディスク制御ユニットDA04
一体型コントローラ

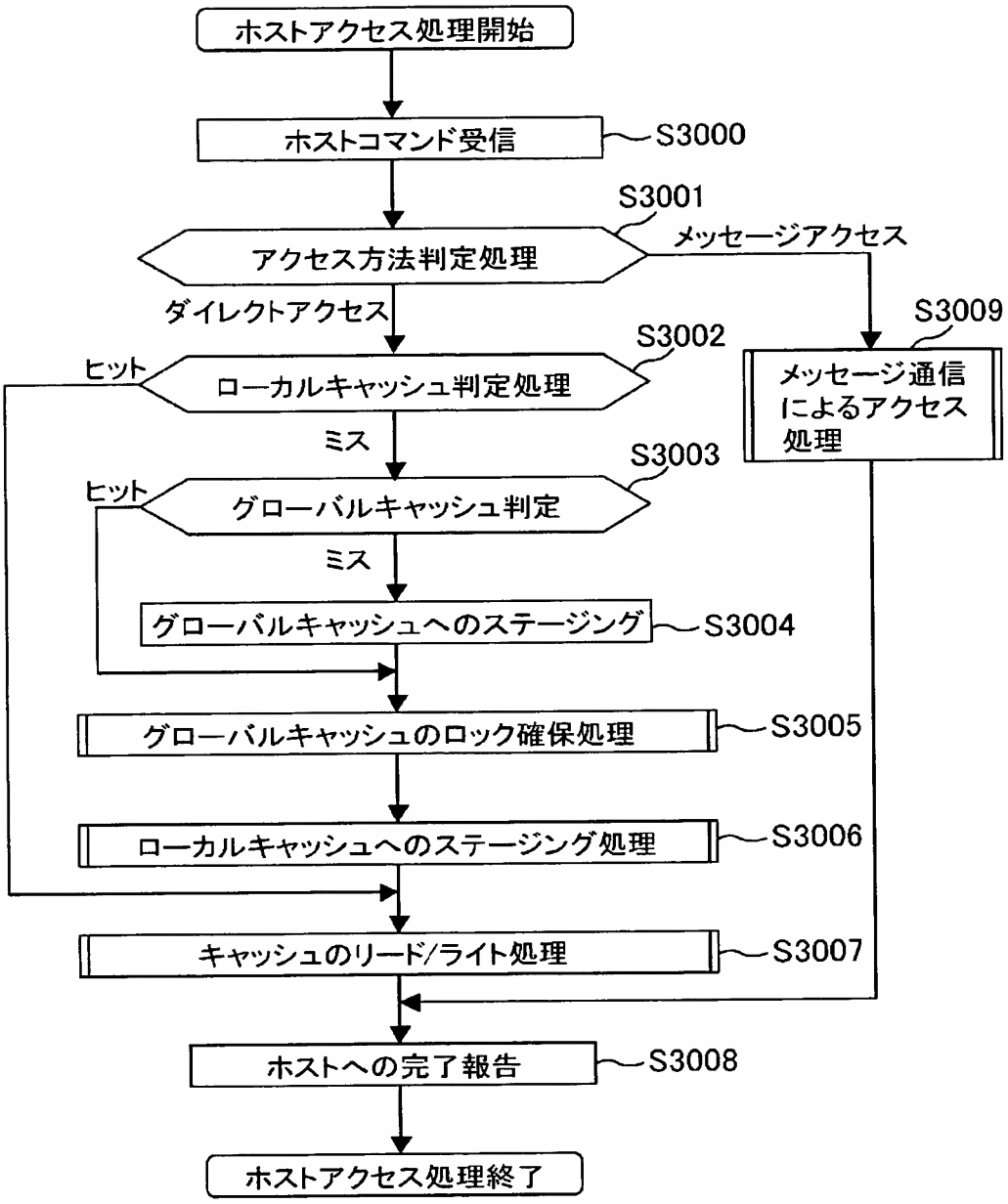
【図２６】


ボリウム管理テーブル移行処理開始
グローバルキャッシュにボリウム管理テーブル領域確保
S1000
ローカルキャッシュのボリウム管理テーブルをロック
S1001
ローカルキャッシュのボリウム管理テーブルを
グローバルキャッシュにコピー
S1002
ローカルキャッシュのボリウム管理テーブルをロック解放
S1003
ボリウム管理テーブル移行処理終了

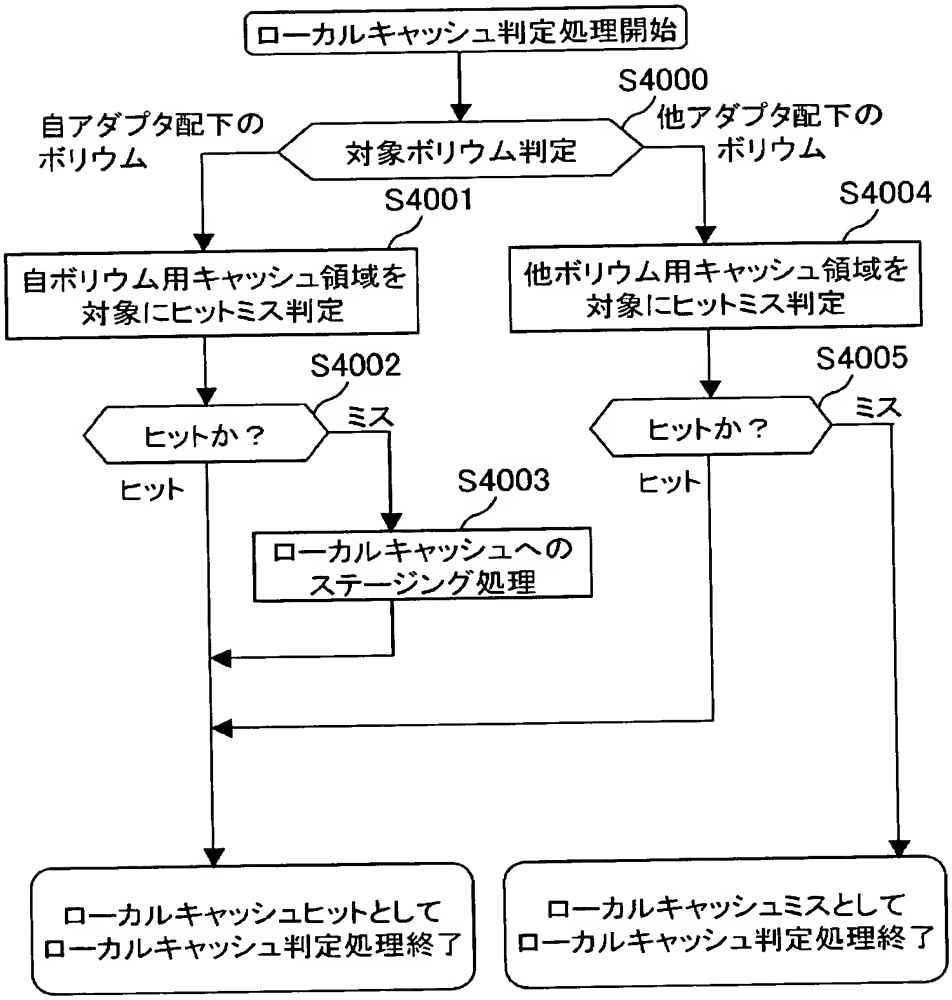
【図２７】


ボリウム作成処理開始
No
アダプタ判定（ストレージアダプタか）
S2000
Yes
ローカルキャッシュのボリウム管理テーブルをロック
S2001
ローカルキャッシュのボリウム管理テーブルに
作成するボリウム情報を追加
S2002
ローカルキャッシュのボリウム管理テーブルをロック解放
S2003
グローバルキャッシュのボリウム管理テーブルをロック
S2004
グローバルキャッシュのボリウム管理テーブルに
作成するボリウム情報を追加
S2005
グローバルキャッシュのボリウム管理テーブルをロック解放
S2006
ボリウム作成処理終了

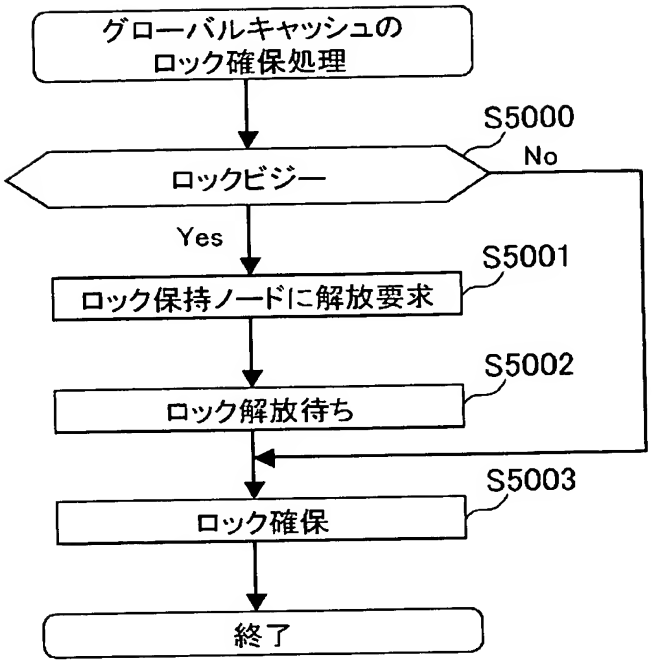
【図２８】


ホストアクセス処理開始
ホストコマンド受信
S3000
S3001
アクセス方法判定処理
メッセージアクセス
ダイレクトアクセス
S3009
S3002
ヒット
ローカルキャッシュ判定処理
メッセージ通信によるアクセス処理
ミス
S3003
ヒット
グローバルキャッシュ判定
ミス
グローバルキャッシュへのステージング
S3004
グローバルキャッシュのロック確保処理
S3005
ローカルキャッシュへのステージング処理
S3006
キャッシュのリード/ライト処理
S3007
ホストへの完了報告
S3008
ホストアクセス処理終了

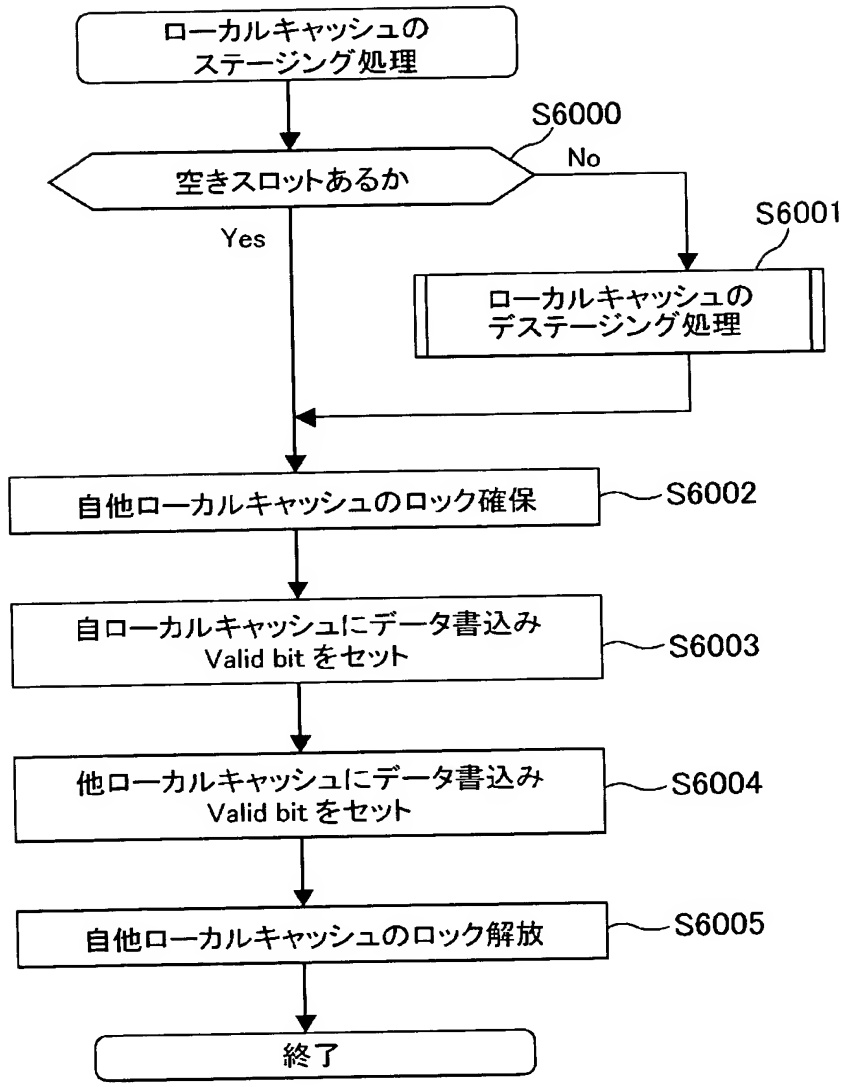
【図２９】


ローカルキャッシュ判定処理開始
S4000
対象ボリウム判定
自アダプタ配下のボリウム
他アダプタ配下のボリウム
S4001
自ボリウム用キャッシュ領域を対象にヒットミス判定
S4004
他ボリウム用キャッシュ領域を対象にヒットミス判定
S4002
ヒットか？
ミス
ヒット
S4003
ローカルキャッシュへのステージング処理
S4005
ヒットか？
ミス
ヒット
ローカルキャッシュヒットとしてローカルキャッシュ判定処理終了
ローカルキャッシュミスとしてローカルキャッシュ判定処理終了

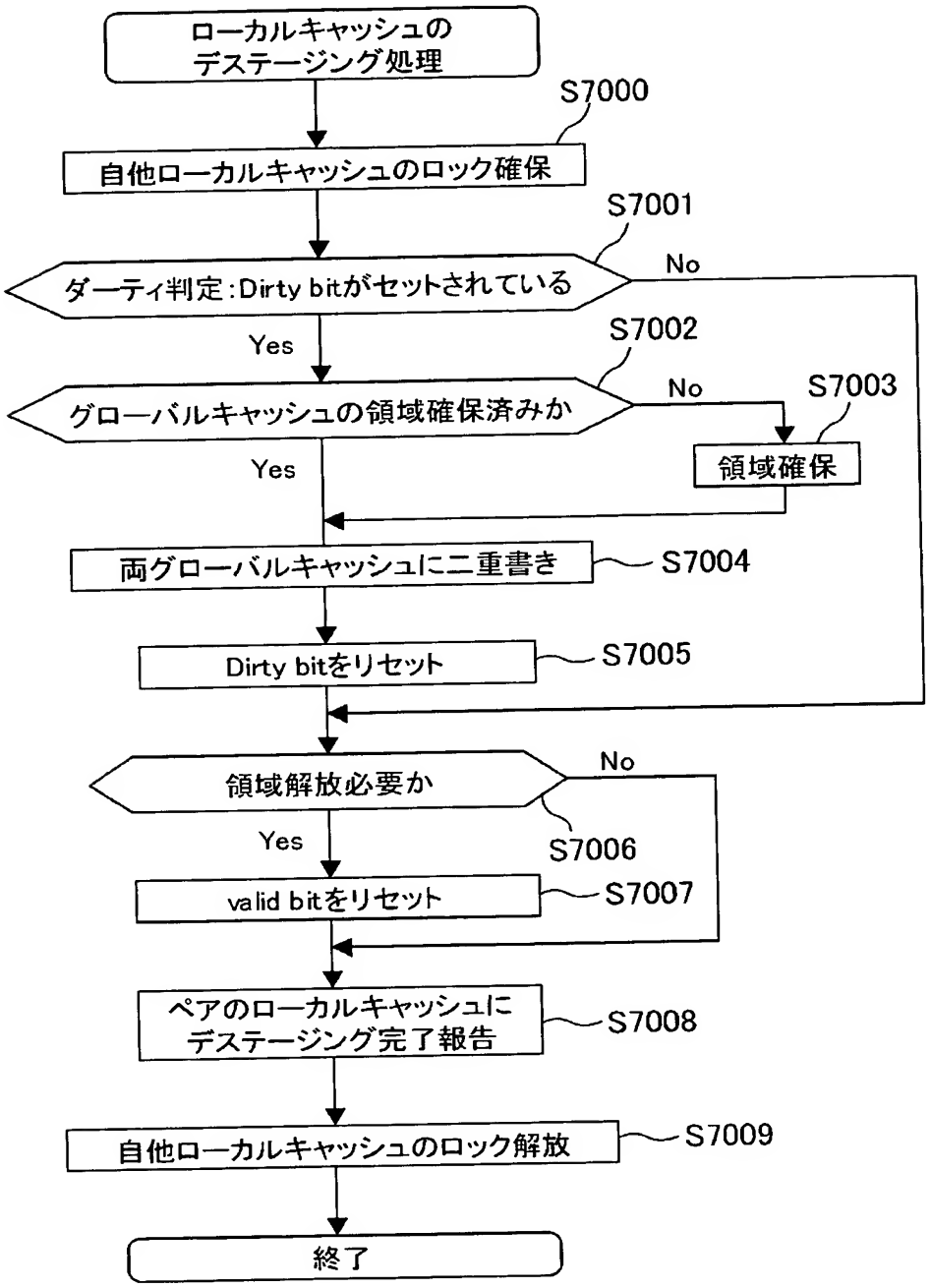
【図３０】


グローバルキャッシュの
ロック確保処理
S5000
No
ロックビジー
Yes
S5001
ロック保持ノードに解放要求
S5002
ロック解放待ち
S5003
ロック確保
終了

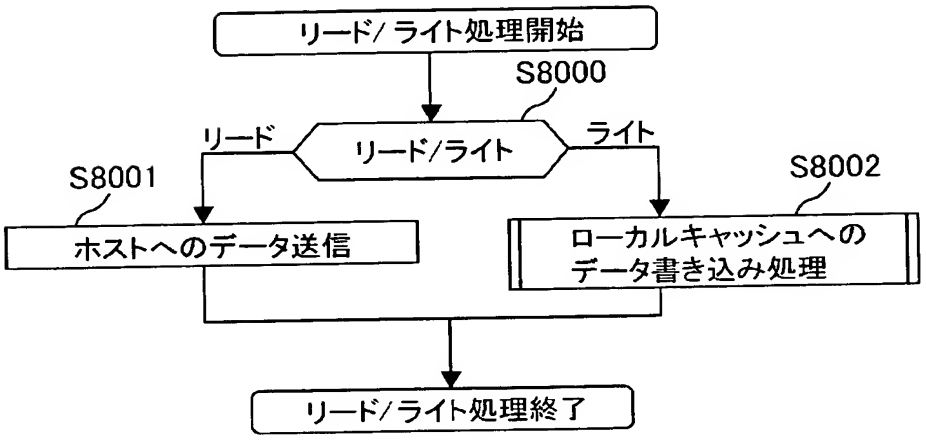
【図３１】


ローカルキャッシュの
ステージング処理
S6000
空きスロットあるか
No
Yes
S6001
ローカルキャッシュの
デステージング処理
自他ローカルキャッシュのロック確保
S6002
自ローカルキャッシュにデータ書込み
Valid bit をセット
S6003
他ローカルキャッシュにデータ書込み
Valid bit をセット
S6004
自他ローカルキャッシュのロック解放
S6005
終了

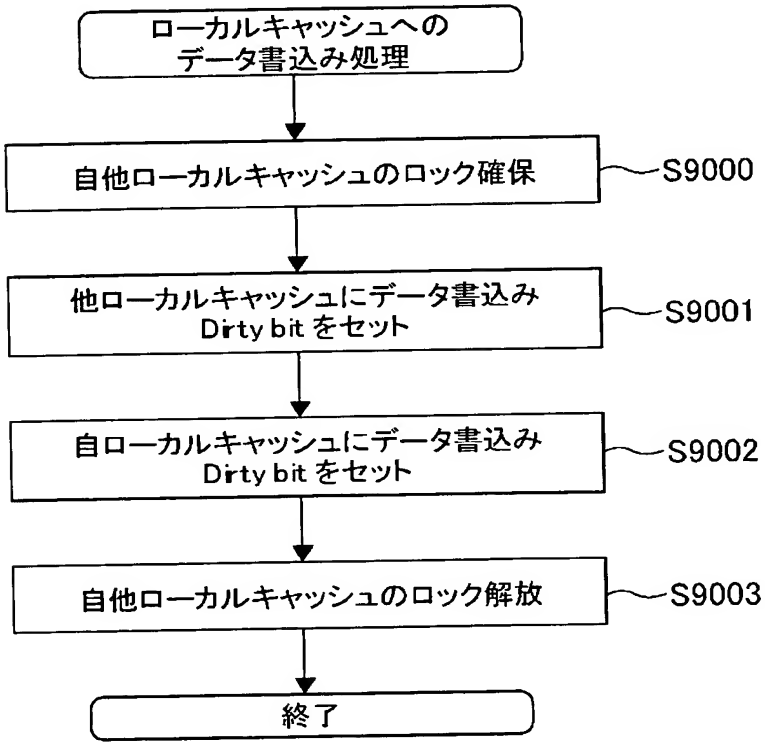
【図３２】


ローカルキャッシュの
デステージング処理
S7000
自他ローカルキャッシュのロック確保
S7001
ダーティ判定:Dirty bitがセットされている
No
Yes
S7002
グローバルキャッシュの領域確保済みか
No
S7003
領域確保
Yes
両グローバルキャッシュに二重書き
S7004
Dirty bitをリセット
S7005
領域解放必要か
No
Yes
S7006
valid bitをリセット
S7007
ペアのローカルキャッシュに
デステージング完了報告
S7008
自他ローカルキャッシュのロック解放
S7009
終了

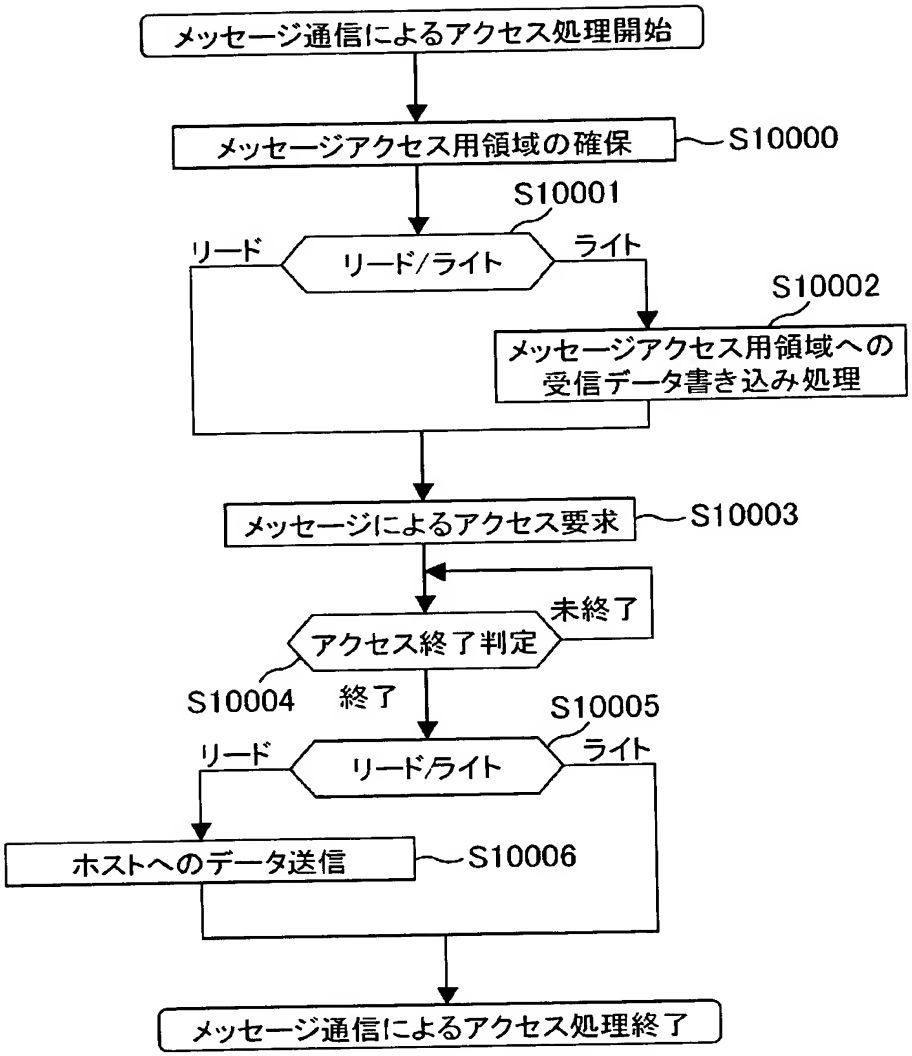
【図３３】


リード/ ライト処理開始
S8000
リード/ライト
リード
ライト
S8001
S8002
ホストへのデータ送信
ローカルキャッシュへの
データ書き込み処理
リード/ ライト処理終了

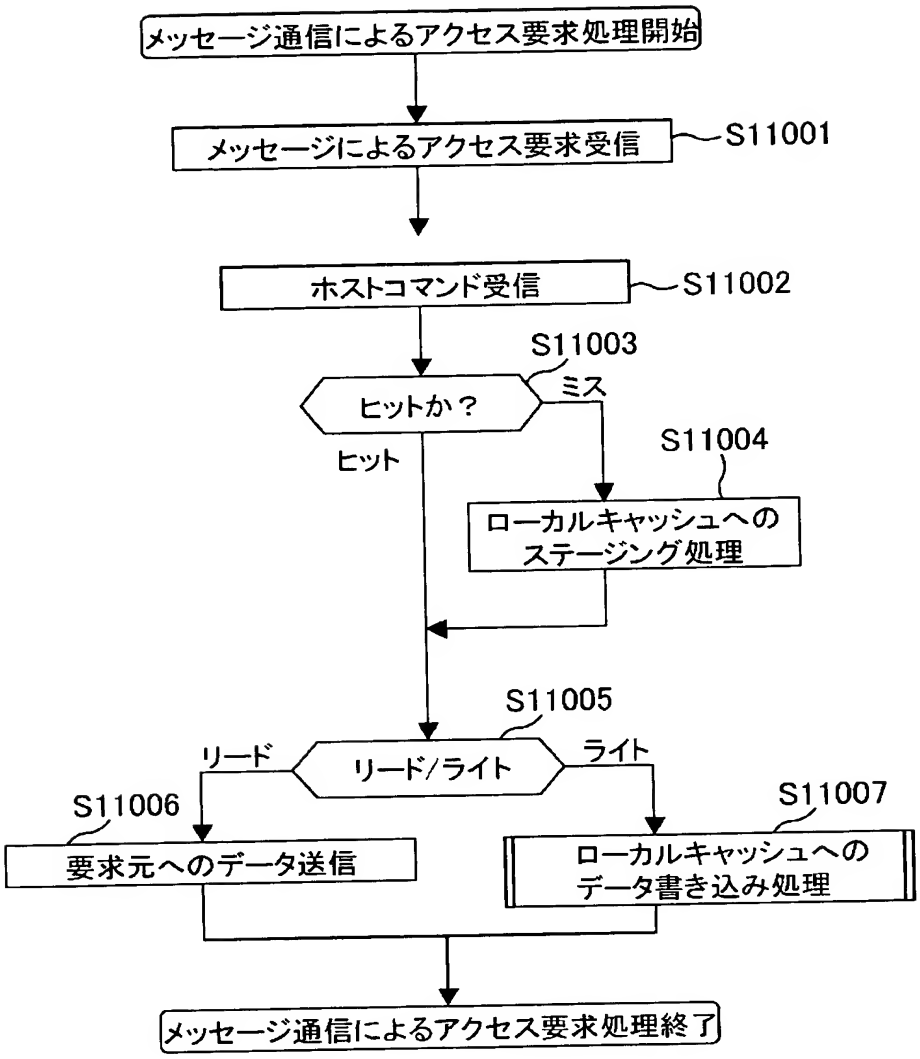
【図３４】


ローカルキャッシュへの
データ書込み処理
自他ローカルキャッシュのロック確保
S9000
他ローカルキャッシュにデータ書込み
Dirty bit をセット
S9001
自ローカルキャッシュにデータ書込み
Dirty bit をセット
S9002
自他ローカルキャッシュのロック解放
S9003
終了

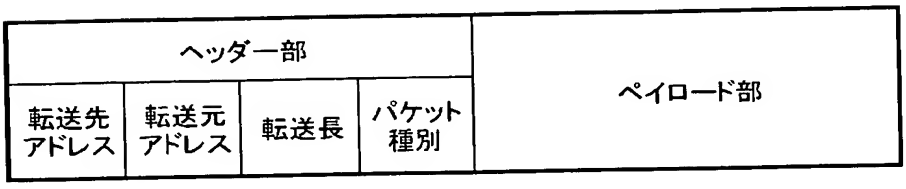
【図３５】


メッセージ通信によるアクセス処理開始
メッセージアクセス用領域の確保
S10000
S10001
リード
リード/ライト
ライト
S10002
メッセージアクセス用領域への
受信データ書き込み処理
メッセージによるアクセス要求
S10003
未終了
アクセス終了判定
S10004
終了
S10005
リード
リード/ライト
ライト
ホストへのデータ送信
S10006
メッセージ通信によるアクセス処理終了

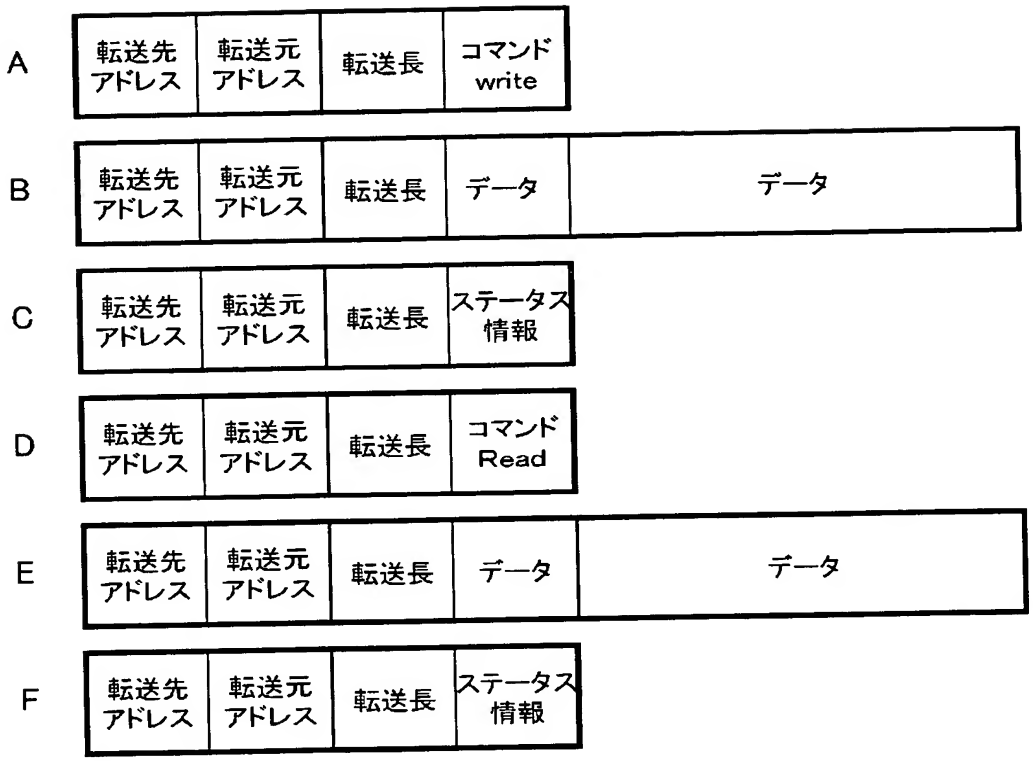
【図３６】


メッセージ通信によるアクセス要求処理開始
メッセージによるアクセス要求受信
S11001
ホストコマンド受信
S11002
S11003
ヒットか？
ミス
ヒット
S11004
ローカルキャッシュへの
ステージング処理
S11005
リード/ライト
リード
ライト
S11006
要求元へのデータ送信
S11007
ローカルキャッシュへの
データ書き込み処理
メッセージ通信によるアクセス要求処理終了

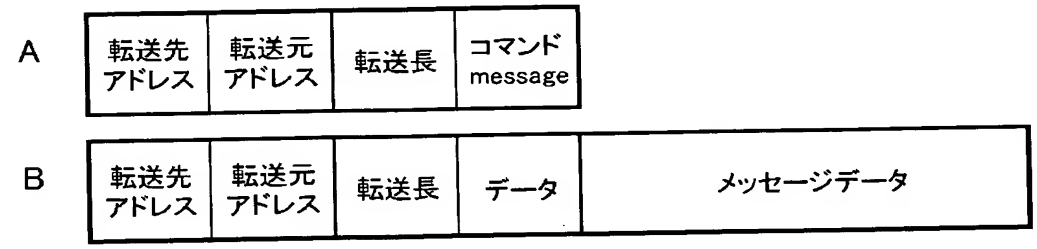
【図３７】


ヘッダー部
転送先アドレス
転送元アドレス
転送長
パケット種別
ペイロード部

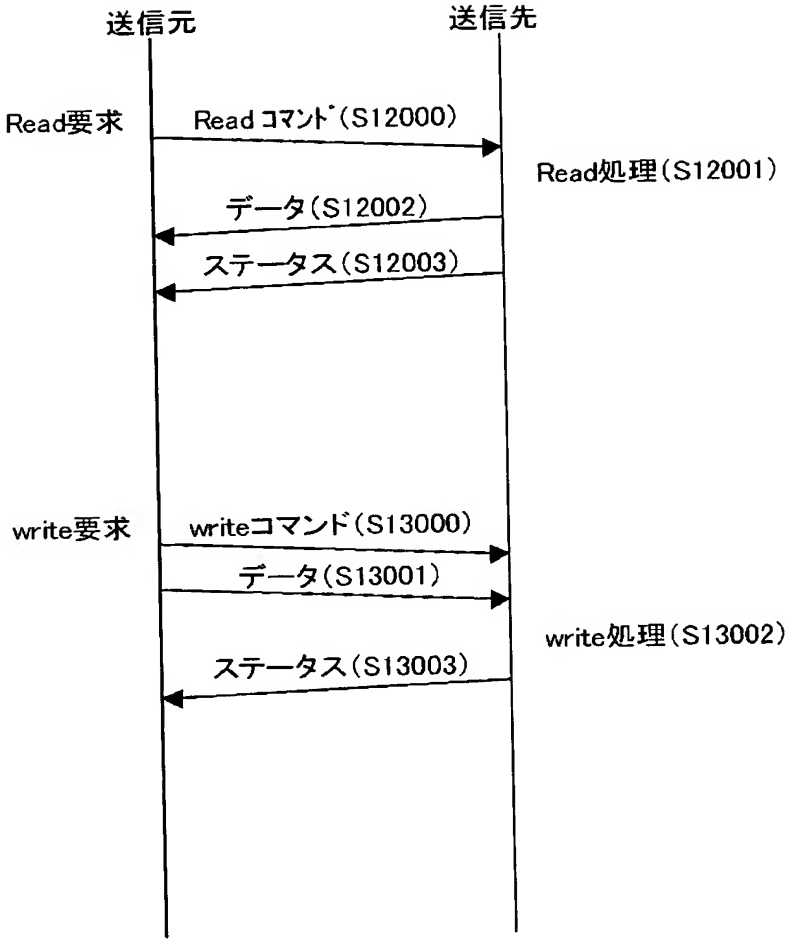
【図３８】


A
転送先アドレス
転送元アドレス
転送長
コマンド write
B
転送先アドレス
転送元アドレス
転送長
データ
データ
C
転送先アドレス
転送元アドレス
転送長
ステータス情報
D
転送先アドレス
転送元アドレス
転送長
コマンド Read
E
転送先アドレス
転送元アドレス
転送長
データ
データ
F
転送先アドレス
転送元アドレス
転送長
ステータス情報

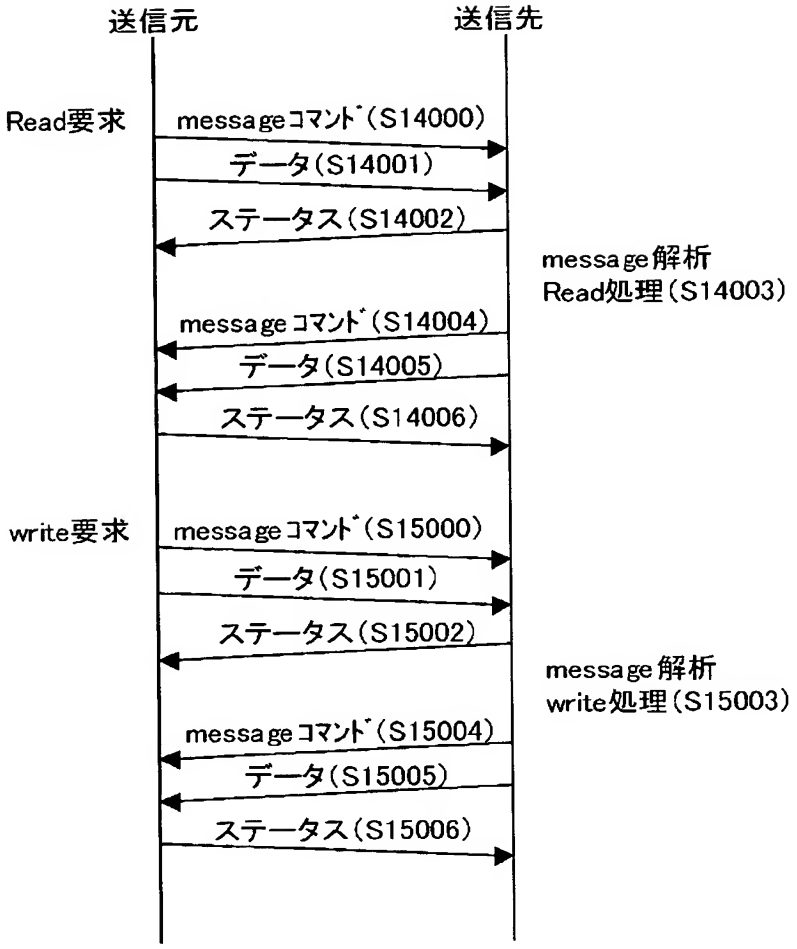
【図３９】


A
転送先アドレス
転送元アドレス
転送長
コマンド message
B
転送先アドレス
転送元アドレス
転送長
データ
メッセージデータ

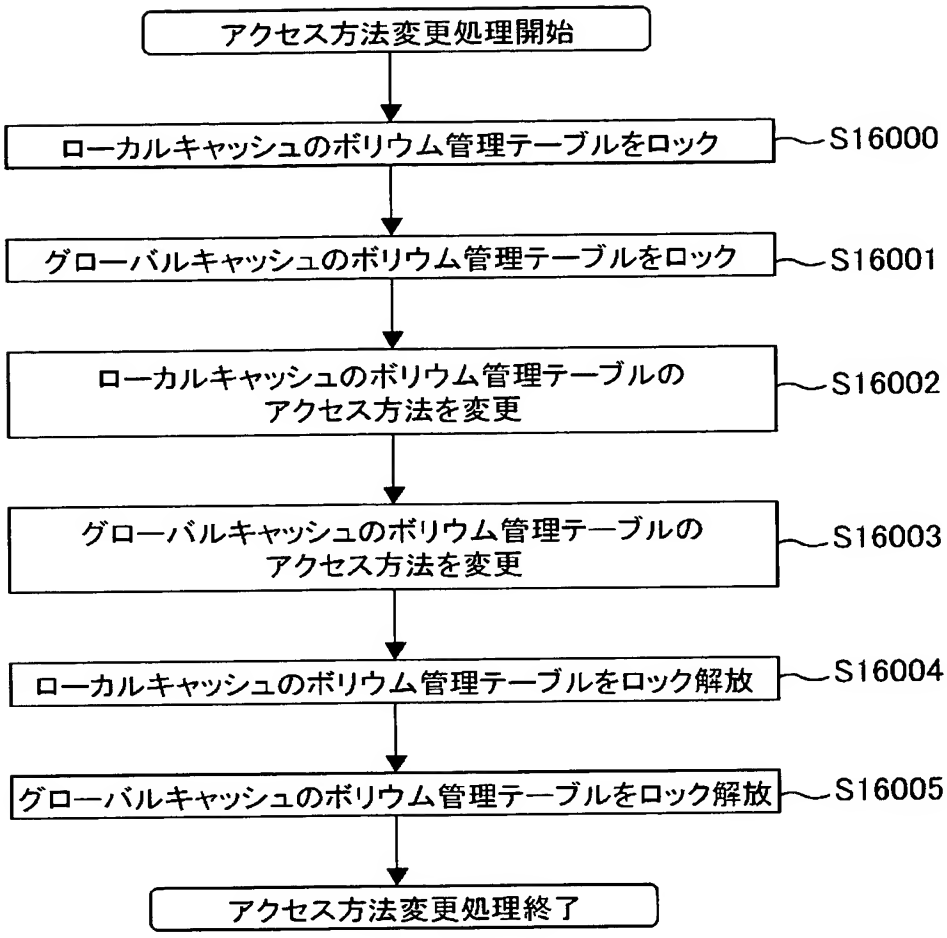
【図４０】


送信元
送信先
Read要求
Readコマンド(S12000)
Read処理(S12001)
データ(S12002)
ステータス(S12003)
write要求
writeコマンド(S13000)
データ(S13001)
write処理(S13002)
ステータス(S13003)

【図４１】


送信元
送信先
Read要求
messageコマンド（S14000）
データ（S14001）
ステータス（S14002）
message解析
Read処理（S14003）
messageコマンド（S14004）
データ（S14005）
ステータス（S14006）
write要求
messageコマンド（S15000）
データ（S15001）
ステータス（S15002）
message解析
write処理（S15003）
messageコマンド（S15004）
データ（S15005）
ステータス（S15006）

【図４２】


アクセス方法変更処理開始
ローカルキャッシュのボリウム管理テーブルをロック
S16000
グローバルキャッシュのボリウム管理テーブルをロック
S16001
ローカルキャッシュのボリウム管理テーブルの
アクセス方法を変更
S16002
グローバルキャッシュのボリウム管理テーブルの
アクセス方法を変更
S16003
ローカルキャッシュのボリウム管理テーブルをロック解放
S16004
グローバルキャッシュのボリウム管理テーブルをロック解放
S16005
アクセス方法変更処理終了

【書類名】　要約書

【要約】

【解決手段】　データ入出力要求を受信するためのホストインタフェース制御部が形成されたチャネル制御ユニットと、前記データ入出力要求に応じて、データを記憶するための記憶ボリウムに対する前記データの入出力制御を行うためのディスクインタフェース制御部が形成されたディスク制御ユニットと、前記データを記憶するためのメモリが形成されたキャッシュメモリユニットと、前記ホストインタフェース制御部と前記ディスクインタフェース制御部と前記メモリとが形成されたストレージ制御ユニットとを挿抜可能な装着部と、前記チャネル制御ユニット、前記ディスク制御ユニット、前記キャッシュメモリユニット、及び前記ストレージ制御ユニットを通信可能に接続する内部接続部とを備えることを特徴とする記憶デバイス制御装置に関する。

【選択図】　図２

特願２００３－１１１４０５

出　願　人　履　歴　情　報

識別番号　　［０００００５１０８］

１．変更年月日　　１９９０年　８月３１日
［変更理由］　　新規登録
住　所　　東京都千代田区神田駿河台４丁目６番地
氏　名　　株式会社日立製作所